Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Accuracy in Reconstructing the Arrangement of Elements Generating Kinetic Depth Displays: B. J. WHITE AND G. E. MUESER 1

Positive and Negative Knowledge of Results on a Pressey-Type Punchboard: W. KAESS AND D. ZEAMAN 12

Remembering the Present States of a Number of Variables: D. B. YNTEMA AND G. E. MUESER 18

Stimulus-Response Generalization with Discrete Response Choice: G. LEVINE 23

The Influence of the Ready Signal and Unpaired UCS Presentations on Eyelid Conditioning: W. R. MCALLISTER AND D. E. MCALLISTER 30

Reversal Learning under Single Stimulus Presentation: D. BIRCH, J. R. ISON, AND S. E. SPERLING 36

Successive Acquisitions and Extinctions as Related to Percentage of Reinforcement: G. D. JENSEN AND J. W. COTTON 41

Effect of Sequential Patterns upon Serial-Position Errors and Acquisition of a Verbal Maze: G. NAMIKAS AND W. J. BROGDEN 50

The Role of Visual Cues in the Performance of Ergographic Work: L. E. JARRARD 57

Replication Report: Alternation following Exposure without Choice: W. N. DEMBER 64

American Psychological Association

Vol. 60, No. 1 July 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 60, No. 1 JULY 1960

ACCURACY IN RECONSTRUCTING THE ARRANGEMENT OF ELEMENTS GENERATING KINETIC DEPTH DISPLAYS[1]

BENJAMIN W. WHITE AND GAYLE E. MUESER

Lincoln Laboratory, Massachusetts Institute of Technology

There is a class of two-dimensional visual displays which appear three-dimensional when, and only when, they are in motion. Displays of this type have been known for some time and have been produced in various ways. Musatti (1924) employed ellipses drawn on a wheel rotated about the line of sight to produce the impression of a round disc tipped out of the plane of the rotating wheel. Similar figures have been studied in detail by Wallach, Weisz, and Adams (1956). This three-dimensional effect, termed by Musatti the "stereokinetic phenomenon," is noted only after an exposure of several seconds, some *S*s achieving it only after *E* has given an appropriate suggestion. A similar phenomenon, achieved with patterns of nonconcentric circles on a turntable rotated about the line of sight, is called Benussi's effect (1927, p. 216). The circles appear to spread apart in depth so that they form a cone rising out of the plane of the turntable, the top of the cone performing a slight circular motion.

A second technique for creating such displays involves the use of a shadow caster such as Gibson's (1957). Various objects such as wire forms or solid objects may be rotated or moved so their shadows are thrown on the translucent screen (Metzger, 1934; Wallach & O'Connell, 1953), or opaque plane forms pasted on transparent sheets may be moved back and forth or rotated so that the changing shadows of these forms are cast on the screen (Gibson & Gibson, 1957).

A third method of producing such displays has involved the use of a cathode-ray tube. Philip and Fisichelli (1945) used such a device to generate rotating Lissajou figures in order to determine the effect of various stimulus parameters on the frequency of perceived reversal of rotation. More recently Green (1958) has employed a high-speed, general-purpose digital computer equipped with a cathode-ray tube output and camera to generate filmstrips of rigid constellations of dots or lines rotating or tumbling.

The stimulus conditions which are necessary and sufficient for producing this perception of depth from movement in a plane display have not been isolated.

[1] The work reported in this paper was performed by Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the United States Army, Navy, and Air Force.

Wallach and O'Connell (1953) and Wallach, Weisz, and Adams (1956) feel that true kinetic depth effect can be obtained only when the display contains lines which, in the course of their movement, change both in direction and in length. Lacking this condition, these authors state, *S*s will not immediately and spontaneously report an impression of a rigid object moving in depth. Stable depth impressions may be obtained when these conditions are not present however. An arrangement such as Metzger used (1934, p. 195) can yield a quite immediate and stable impression of a rigid arrangement of elements turning in depth, and in such a display the lines change neither in length or direction. Gibson (1957) and Gibson, Gibson, Smith, and Flock (1959) express the opinion that the impression of rigid motion in depth requires a display which is a perspective transformation. It is clear from Green's (1958) displays that a perspective transformation is not a sufficient condition to generate such an impression, and displays such as those used by Metzger with parallel projection and with asymmetrical arrangements of elements indicate that projective transformations are not a necessary condition. Green's experiments indicate that the strength of the depth effect is strongly influenced by the number of elements in the display and by their arrangement.

The accuracy of judgments of angle or tilt which can be made from such moving two-dimensional displays has been investigated both by Wallach and O'Connell (1953), Gibson and Gibson (1957), and Gibson et al. (1959). The *S*s appear to underestimate slants which deviate only slightly from frontal parallel, and are more accurate in their judgments of greater slants. Gibson and Gibson (1957) found that the accuracy of judged slant was little affected by the regularity of the form or pattern with a continuous perspective transformation.

The experiments reported here were designed to explore further the factors which affect the accuracy with which information about the arrangement of objects in depth can be transmitted to human observers by means of such moving displays.

General Method

Apparatus.—The apparatus, which is similar to that used by Metzger (1934, p. 195) is diagrammed in Fig. 1. The light source was a 500-w. projector (L) with a .32-in. diameter aperture (D) placed in front of the lens. At a distance 21 ft. in front of the projector was a turntable (T) whose speed was continuously variable up to 100 rpm. On the turntable was mounted a slab of Plexiglas drilled with a rectangular matrix of holes ¾ in. on centers. Pegs of various diameters and shapes could be inserted into these holes. In front of the turntable and set into the wall of the room was a flashed opal screen 17¾ in. wide and 1½ in. high (W), whose center was in line with the turntable's center of rotation and with the light source. When pegs were placed in the matrix and the light source turned on, the vertical shadows of the pegs fell on this screen. Rotation of the turntable caused the peg shadows to move from side to side. The geometry of the arrangement of light source, turntable, and screen made the projection essentially parallel which meant that the thickness of the cast shadows did not vary detectably during rotation. Preliminary tests showed that reversals of perceived direction occurred both with parallel and with polar projection. Momentary views of typical displays are shown in Fig. 2. The *E* sat beside the turntable near controls for the light source and turntable. The *S* sat in semidarkness at a table in an adjoining room. His eyes were at a level with and 27 in. from the display screen. A dim light was placed on the table so that *S* could see to record his responses.

Subjects and schedule of experiments.—Nine staff members of Lincoln Laboratory, four

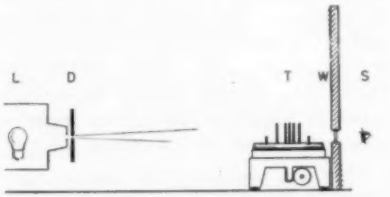

Fig. 1. Schematic view of the display generating apparatus. The component parts are described in the text.

men and five women, served over a period of several months as *Ss* for these experiments. The *Ss* were run individually in sessions that lasted between 20 and 45 min. They were not given any information about the accuracy of their performance on any session.

Before presenting each display, *E* placed a predetermined arrangement of pegs in the plastic matrix on the turntable, set the speed of rotation, and the exposure duration. After a ready signal, *E* started the turntable and turned on the light source. With five-element displays, each *S* was first shown a standard pattern, an H, under the conditions which were to obtain for that session. This was followed by 2 practice displays and then 8 or 10 displays, the responses to which were scored. In the experiment with only two elements, *E* described the display to *Ss*, let them observe the display for a short time, and then proceeded into the experiment. The order of experiments was the same for all *Ss*, but within a given experiment conditions were balanced between *Ss*.

Two-Element Moving Displays

Method

Throughout these experiments two pegs were mounted symmetrically on either side of the center of rotation so that the display was of two vertical bars moving with equal frequency and amplitude but 180° out of phase. At the point of maximum excursion the bars were 9 in. apart. The turntable speed was set at 40 rpm and the exposure duration at 90 sec. Twice during each revolution of the turntable the two bars met, and for a brief period there was displayed only one dark bar which varied in thickness.

The simplicity of the two-bar display, and the fact that it gives rise to an unstable KDE, make it a promising situation for studying factors that inhibit or enhance KDE. The two-bar display has several modes of appearance which differ in apparent depth. At one moment it may be seen as two bars fixed on some unseen disc which is rotating at constant speed with direction of rotation changing from time to time. Indeed, reversals may become so frequent that the unseen disc may appear to pivot each time the two bars coincide at the center of the display. Regardless of the reversals in direction, the display appears definitely three-dimensional with the bars rotating in depth. In a two-dimensional mode of appearance, the bars may seem to meet at the center and recoil, or "bounce." The bars do not cross, but collide on a common path. There is a third, less frequently observed mode of appearance which lies between these two in its apparent depth. In this, the bars appear to be traveling in two straight parallel paths, one behind the other. They merely cross and recross, each bar in its separate path. The paths are straight, not curved in depth, as in the first mode of appearance. They have no point in common as in the second mode of appearance.

The *Ss* were shown a two-bar display such as A or B in Fig. 2 with instructions to report various modes of appearance as they occurred. All *Ss* reported both the circling and the bouncing modes. They were then instructed to depress a response key on subsequent presentations whenever the bars appeared to be bouncing. This key activated a clock which cumulated the number of seconds the display appeared in this mode, and a counter which cumulated the number of perceived reversals. Each *S* viewed for 90 sec. the two-bar display under 12 conditions. The conditions were formed by combinations of: two directions of movement, horizontal and vertical; two kinds of peg shapes, identical and dissimilar (A and B in Fig. 2); and three fixation points (a, b, c or c′ in Fig. 2A).

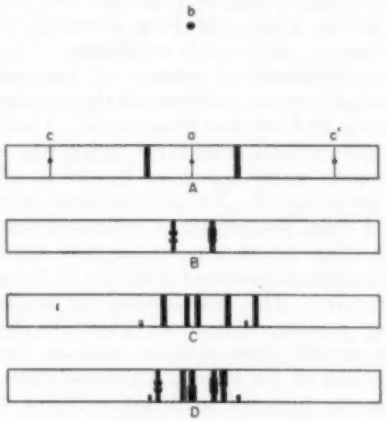

Fig. 2. Momentary views of four types of moving display employed in the experiments. (A) Two elements identical in shape, with alternate fixation points shown at a, b, c, and c′. (B) Two elements discriminably different in shape. (C) Five elements identical in shape, and two reference markers. (D) Five elements discriminably different in shape, and two reference markers.

TABLE 1

MEAN TIME IN SECONDS THE TWO-ELEMENT DISPLAY APPEARED TWO-DIMENSIONAL
(Each mean is based upon nine observations. The exposure duration was 90 sec.)

Fixation Point	Elements Identical		Elements Discriminably Different	
	Horizontal Movement	Vertical Movement	Horizontal Movement	Vertical Movement
Central	38.00	44.89	26.22	25.11
Above	18.00	17.22	17.22	12.67
Lateral	4.22	13.78	1.67	15.11

Using only his preferred eye, *S* viewed the display through a dove prism that was mounted so that the image of the display screen could be rotated 90°. Lateral fixation was to the left of the center of the display for the left-eyed *S*s and to the right for the right-eyed *S*s, an arrangement which kept the image of the central part of the display from falling on the blind spot.

Results

Table 1 gives the mean number of seconds, averaged over the nine *S*s, during which the bars appeared to bounce under each condition. This two-dimensional aspect of two-peg display was in evidence on the average only 19.5 sec. in a 90-sec. trial. There were several conditions under which one or more *S*s saw no bounce at all during a trial. An analysis of variance of these bounce times indicated that fixation point was by far the most important source of variance in these results. The two-dimensional mode or bounce was most likely to be reported when fixation was at the center of the display and least likely to be reported at Fixation Point c or c′. There was a significant interaction between fixation point and direction of movement. Central fixation yielded the greatest bounce time under both directions of movement. The other fixation points differed only with horizontal movement. Direction of movement alone, however, did not affect amount of bounce time significantly. There were large observer differences that interacted with direction (some *S*s reporting more bounce with vertical motion, and others reporting more with horizontal motion).

Making the pegs of unlike shape reduced the amount of bounce time reported, especially with central fixation where one would expect the influence of shape to be most marked. When *S*s saw the differently shaped bars as bouncing, there was a peculiar interchange of parts at the moment of collision. Had the shape difference been made more pronounced, reports of bounce would probably have been less frequent. The fact remains, however, that the two-dimensional mode of appearance is seen in a two-element display even when it necessitates an abrupt change in shape of each bar.

FIVE-ELEMENT MOVING DISPLAYS

Method

Stimulus patterns.—The five-element patterns used in these experiments were a restricted subset of all possible arrangements of five pegs in a 4 × 5 matrix. Arrangements were eliminated which contained three or more pegs in a straight line, or which had four pegs forming two parallel lines, or which did not have at least one peg in Rows 1 and 5 and one peg in Columns 1 and 4. In all of the experiments with five-element displays the patterns used were selected from the 588 arrangements that met these specifications. The *S* was not informed about these restrictions, but was told that five pegs arranged in a matrix of four columns and five rows cast the shadows on the screen. He was supplied with a response sheet on which 4 × 5 matrices were printed. After studying the display, he marked on the response sheet the five matrix positions that he thought had been occupied by pegs on that trial.

Such a set of pegs gave a display on a light background of five identical dark lines, moving from side to side sinusoidally with the same frequency but with differing amplitudes and phases (see Fig. 2C). The velocity of a particular dark bar was maximal as it crossed the center of the screen and ap-

proached zero as it reached the limit of its right or left excursion and changed direction. This velocity is given by $2\pi rS$/min. where r is the distance of the peg from the center of rotation, and S is the speed of the turntable in rpm. At a given moment, all pegs between the light source and the center of rotation cast shadows which were moving in one direction, and all pegs between the screen and the center of rotation cast shadows which were moving in the opposite direction. Since the projector provided a virtually parallel light source, there was no perceptible change in the thickness of a shadow as the peg changed position. Thus a dark bar moving from left to right on the screen could be the shadow of a near peg rotating in a counterclockwise direction, or of a far peg turning clockwise. There was nothing in the display to enable *S* to discriminate between these two possibilities.

Scoring procedure.—In order to understand the rationale of the scoring procedure, an analysis of these moving displays is in order. Suppose that pegs are placed in the matrix to form the letter "p" and the turntable is rotated clockwise. From *S*'s standpoint there is nothing in the display to indicate the direction of rotation, and as a matter of fact different *S*s report opposite directions, and a given *S* will experience reversals in direction while observing the display. Should *S* report the direction of rotation to be clockwise, he will also report the figure to be a "p," or its 180° transformation, a "d." Since the whole arrangement is rotating, there is no a priori reason for declaring one of these identifications to be more correct than the other. If, on the other hand, *S* sees the direction of rotation as counterclockwise, he will report the figure "p" as its mirror image, namely a "q" or a 180° rotation of this mirror image, a "b." Thus it is clear that there are four transformations of any given arrangement, any one of which may be reported by *S*. There is nothing in the moving display to permit *S* to make any discrimination between the four.

For each pattern employed in an experiment a mask was cut on a card so that a hole was punched for every matrix position occupied by a peg on that trial. This mask was placed over the response matrix and a count made of the marked positions. For the reasons outlined above, the mask was put in all four possible orientations over the response matrix and a count made for each position. The score was defined as the highest of these four counts. Scores could take any integer value from zero to five.

Results

Peg thickness and spacing.—The independent variables in this experiment were the thickness of the pegs (2-, 3-, 4-, and 5/16 in.) and the spacing of the holes in the matrix (3- 6-, and 9/8 in. on centers). The 5/16-in. peg was used only on the smallest matrix; otherwise, all peg sizes were employed on all three matrices, making a total of ten conditions. For all conditions the exposure time was 30 sec. and the turntable revolved at 8 rpm.

The major finding of this experiment was that six of the *S*s did not score significantly above chance expectancy. An a posteriori chance score was obtained by scoring a sample of 40 responses of each *S* with a mask randomly selected from a set which did not include the pattern displayed. The mean chance and obtained scores are shown in Table 2. Tukey's signed-rank test indicated that six of the *S*s had scores which were indistinguishable from the chance distribution ($P > .05$) and three of the *S*s had scores which were clearly above the chance distribution ($P < .001$).

Results of a separate analysis of the scores of the three *S*s whose scores were above chance showed significant individual differences, but no significant differences attributable to peg thickness or spacing. There was no significant practice effect, either from condition to condition in the course of the experiment, or from trial to trial within a given condition.

Even though his scores were below chance expectancy *S* might still have an accurate percept of certain topological characteristics of the arrangements. To check this possibility, one additional scoring procedure was tried. Throughout this experiment half the patterns used for each condi-

TABLE 2

MEAN NUMBER OF ELEMENTS CORRECTLY PLACED, AND RANGE OF INDIVIDUAL SUBJECT MEANS FOR FIVE-ELEMENT DISPLAYS VIEWED UNDER VARIOUS EXPERIMENTAL CONDITIONS

Experimental Condition	Grand Mean	Range of Individual *S* Means
Peg thickness and spacing (Mean of 10 conditions)	2.55	2.09–3.38
(A posteriori chance score)	(2.25)	(1.90–2.42)
Reference markers	3.33	2.25–4.25
Speed of rotation:		
1 rpm	2.84	2.00–4.00
2 rpm	3.39	2.25–4.38
8 rpm	3.27	2.25–4.25
32 rpm	3.25	1.88–4.50
Exposure duration:		
7.5 sec.	3.16	2.40–3.60
15 sec.	3.20	2.50–4.00
30 sec.	3.74	2.50–4.90
60 sec.	4.27	2.70–5.00
Subject control of rotation:	3.28	2.15–4.55
Displacing center of rotation:		
Eccentric	3.25	2.00–4.50
Centered	3.66	2.20–4.30
Direction of movement:		
Horizontal	3.78	2.60–4.90
Vertical	3.51	1.80–4.30
Oblique (45° or 135°)	3.14	2.30–4.90
Discriminability of peg silhouettes:		
Shapes discriminably different	3.80	2.25–4.62
Shapes identical	3.28	2.25–4.25
Tachistoscopic presentation:		
1/4-sec. exposure	3.55	3.20–3.90
1-sec. exposure	4.07	3.80–4.50
2-sec. exposure	4.40	3.90–4.80
Binocular regard (30-sec. exposure)	4.59	2.50–4.80
Monocular regard (30-sec. exposure)	2.10	1.60–2.50

tion contained at least one peg whose shadow was always surrounded by the shadows of other pegs. In other words, at no time was its shadow on the periphery of the display. Such arrangements were called "included." The other arrangements used in any trial were all "convex"; they had peg arrangements such that the shadow of every peg at some time during a revolution of the turntable took a peripheral position. Just as the display arrangements could all be classified as "convex" or "included," so also could the response patterns. Thus it was possible to test by means of a χ^2 whether *S*s perceived this topological aspect of the pattern with a frequency above chance expectancy. By this criterion every *S* performed significantly above chance. The percentage of correct responses for all *S*s was 79 (572/720) with individual *S*s ranging from 70% to 92% correct. A two-way analysis of variance on these scores (9 *S*s, 10 conditions of matrix and peg size) showed significant differences between *S*s, but not between conditions. Accuracy in reproducing arrangements in these moving displays is not affected by changes in the spacing and thickness of the moving bars, at least within the limits of this experiment.

Reference markers.—The fact that two-thirds of the *S*s did not score above chance (mask scoring) made it necessary to make the display more readable so that the effect of other experimental manipulations could be evaluated. This was accomplished by inserting two short thin pegs at the head and foot of the matrix to give an indication of the matrix's orientation (see Fig. 1, C and D). The *S*s were told that the short dark bars on the display marked the head and foot of the matrix, so that when the bars were close together in the center of the screen they were looking down the columns of the matrix, and when the bars were maximally far apart, they were looking down the rows.

With these reference markers, each *S* was given eight 5-peg arrangements composed of $\frac{3}{16}$-in. pegs whose positions in a matrix were $\frac{3}{4}$ in. on centers, revolving at a speed of 8 rpm for an exposure time of 30 sec. Except for the presence of the reference markers this was a replication of one of the 10 conditions used in the previous experiment. With the reference markers the mean score was 3.33 as compared to a mean of 2.53 on the condition without markers. This highly significant mean differ-

ence ($P < .01$) indicates that the addition of reference markers substantially increased the accuracy with which *S*s could reproduce the five-peg arrangements. Significant individual differences remain, but no significant $S \times$ Condition interaction was found. Since the reference markers did raise the scores above chance level, they were employed in all subsequent experiments.

Speed of rotation.—At very high speeds of rotation the dark bars in the display move too fast to be discriminated as individual bars. When the display is static, there is no impression of depth. It is not obvious whether between these extremes there is a fairly narrow range of speeds within which *S*s can accurately describe the arrangement of elements rotating in depth. The purpose of this experiment was to study the effects of a wide range of speeds on the accuracy of reproducing the display. The conditions were exactly the same as for the previous experiment: 30-sec. exposure time and $\frac{3}{16}$-in. pegs in a matrix whose positions were $\frac{3}{4}$ in. on centers. The speeds chosen were 1, 2, 8, and 32 rpm.

Scores averaged over eight patterns for each of nine *S*s were 2.84, 3.39, 3.27, and 3.25 for the four speeds, respectively. Analysis of variance disclosed a barely significant difference attributable to speed ($P < .05$.) Application of Duncan's (1955) range test to the data showed the scores at the slowest speed, 1 rpm, to be significantly lower than the scores for speeds of 2 and 8 rpm. Scores for the three fastest speeds (2, 8, and 32 rpm) did not differ significantly from each other. As expected, performance becomes less accurate at fast and slow rates of rotation. As in the previous experiments, differences between *S*s were significant but did not interact with the four speed conditions.

Exposure duration.—Four exposure durations were employed: 7.5, 15, 30, and 60 sec., with conditions of speed, peg spacing, and peg width the same as those employed in the reference marker experiment. With these durations, *S*s saw 1, 2, 4, and 8 complete cycles of the moving pattern.

Mean scores for the four exposure durations were respectively 3.16, 3.20, 3.74, and 4.27 elements correctly located per pattern. An analysis of variance showed the effect of exposure time to be highly significant ($P < .01$). Duncan's range test (1955) indicated that at the .05 level of significance, all the distributions differed from each other except those for the two shorter exposure times. As in previous experiments, there were highly significant differences between *S*s. For example, one *S* on the 60-sec. exposure time responded correctly to every element, while another *S* under the same condition responded correctly to slightly more than half the elements. There was no significant *S* by Condition interaction.

Subject control of turntable rotation.—So that *S* could control the direction and speed of rotation of the display elements, the turntable on on which the peg matrix was placed was modified by the addition of a friction drive, a flexible shaft, and a crank. In all other respects the display conditions were the same as in the reference marker experiment.

The results of this experiment were compared with means for the same eight *S*s computed from speeds 2, 8, and 32 rpm reported in the experiment on speed of rotation. By this comparison, there was no significant advantage to giving *S*s control over the rotation. The mean number of

pegs correctly placed per pattern under the *S*-controlled rotation was 3.28, while that computed from the speed of rotation experiment for speeds 2, 8, and 32 rpm was 3.31. The ubiquitous significant differences between *S*s again were in evidence.

Displacing center of rotation.—For this experiment $\frac{1}{8}$-in. pegs were used on a matrix whose positions were $\frac{3}{8}$ in. on centers. One matrix had its geometric center coincident with the center of rotation, while the other had its center $3\frac{1}{2}$ in. from the center of rotation. Speed of rotation was 8 rpm and exposure time 30 sec.

Phase differences between elements diminish as the center of rotation is displaced from the center of the matrix. Thus, with a centered display, half the bars move in one direction and half in the opposite direction, on the average, in any temporal interval. When the matrix is displaced from the center of rotation, there is an increased likelihood that all the bars will move in the same direction in a given interval.

The mean accuracy score for arrangements coincident with the center of rotation was 3.66, and for the eccentric arrangements 3.25. The difference between the two means was significant at the .05 level. Again *S* differences were significant, and no $S \times$ Condition interaction was found.

Direction of movement.—With his preferred eye, *S* viewed the display through a dove prism so arranged that the image of the display could be rotated a desired number of degrees. Three orientations of the display were used, horizontal, vertical, and oblique (45° for right-eyed and 135° for left-eyed *S*s). Other conditions were the same as those used in the reference marker experiment.

Average number of pegs correctly located per pattern was 3.78 with horizontal movement, 3.51 with oblique movement, and 3.14 with vertical movement. Analysis of variance indicated the difference between these means to be significant at the .05 level. More striking were the differences between *S*s and an interaction between *S* and Direction, significant at the .01 level. Some *S*s were able to reproduce the patterns more accurately when the movement was vertical, while others did better with horizontally moving displays.

Discriminability of peg silhouettes.—The pegs in this experiment were discriminably different in silhouette, shown in Fig. 2D. All were round in cross section so that shadow shape did not change as the matrix rotated. Conditions were otherwise the same as those employed in the reference marker experiment.

With discriminably different elements, average score for eight *S*s was 3.80. In a control condition, using the same arrangements with identically shaped pegs, as in Fig. 2B, *S*s had an average score of 3.28. The difference between the means for the two conditions was highly significant ($P < .01$). The significant differences between *S*s showed no significant interaction with experimental condition.

Five-Element Static Displays

Two experiments were performed in order to determine how well *S*s could reproduce the same five-element patterns from static displays. As far as possible, conditions were kept the same as those which obtained in the moving display experiments so that the performance figures could be compared.

Tachistoscopic presentation.—Forty patterns from the same population

of five-element arrangements employed in the experiments described above, were drawn and photographed to make a filmstrip, one frame of which is shown in Fig. 3. The filmstrip was projected from a device which permitted control of exposure time while *S* sat in semidarkness holding a clipboard with a response sheet. He was shown six practice frames, two at each of the following exposure times: ¼ sec., 1 sec., and 2 sec., with instructions to reproduce the pattern on an answer sheet which was identical to that used in the five-element moving display experiments. Following the practice items, *S* was given 10 items at each of these exposure times, the order of exposure times being counterbalanced between *S*s. Responses were scored as before, except that the mask was placed in only one correct orientation.

Accuracy scores with this method of presentation were comparable to those found in the Exposure Duration experiment described above, although the exposure times in the present experiment were shorter by a large factor than those used in the moving display. Longer exposure times yielded higher scores in this experiment, but no significant differences were found between *S*s.

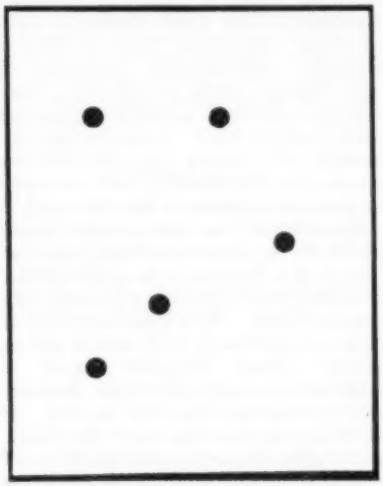

FIG. 3. Typical five-element pattern used for tachistoscopic presentation.

Binocular regard.—The display employed in this experiment was like that in the five-element moving display experiments except for the following modifications: Instead of looking at the moving shadows of the pegs, *S* viewed the pegs directly with his head in a headrest. A light table was placed behind the pegs to give uniform illumination. The pegs were painted a flat black, and thus appeared as dark silhouettes against a light illuminated ground. The display was static. In all other essential respects the display was the same as that used in the reference marked experiment reported above. The *S* was shown a practice item consisting of an H arrangement and a sample response sheet marked to show the locations of the elements in this arrangement. Following this practice item, *S* was shown 10 patterns with an exposure time of 30 sec. He was permitted to write his response during or after the pattern was exposed and was given as much time as he needed to respond.

Under this condition the mean accuracy of reproduction score was 4.59, slightly higher than the mean score obtained with the 60-sec. exposure with the moving display. Since most *S*s had made their responses by the time the display had been exposed for 15 sec., it is doubtful whether a longer exposure time would have resulted in further improvement.

Monocular regard.—The procedure was exactly the same as that used in the previous experiment, except that the nonpreferred eye was oc-

cluded. Under this condition, with little or no information available to *S* about the arrangement of the elements in depth, the average score was 2.10, the lowest average score under any condition.

Discussion

It is an astonishing perceptual feat to be able to look at identically shaped elements moving along identical paths, with identical frequency, and to use only the differences between their phase angles and amplitudes to arrive at a percept of a rigid arrangement of elements rotating in depth. These experiments have shown that under certain conditions such percepts are in good accord with the geometry of the apparatus used to generate the displays. Such performance is possible only if the perceptual system is capable of rather precise temporal and spatial integration of the information coming in during the exposure interval.

This processing of the input is evidently little affected by a number of display characteristics: the speed, direction, and eccentricity of rotation, or the thickness or proximity of the elements in the display. It can be facilitated significantly by providing reference markers, by making the elements discriminably different in shape, and by increasing exposure time.

The most dramatic change in performance was brought about by the introduction of the reference markers. Two-thirds of the *S*s were unable to reproduce the arrangements with an accuracy significantly greater than chance before the markers were introduced. With the markers, all *S*s were able to perform at better than chance levels of accuracy. Though the level of performance was low without the reference markers, it should not be assumed that the display failed to produce an impression of depth. Even though *S*s were not able to reproduce the patterns accurately, they reported immediate and stable impressions of a rigid arrangement rotating in depth. The accuracy performance in all likelihood could have been improved still further by introducing additional reference markers, or by altering their placement in the matrix. Those desiring to use such displays in an applied situation would be well advised to give careful attention to the coordinate or reference system which is employed, since it can materially affect operator accuracy.

Making the elements discriminably different in shape greatly increased the proportion of time the two-element display was seen in its three-dimensional mode, especially with fixation at the point of crossing, and significantly increased the accuracy of performance in reproducing the five-element arrangements. These results are in general accord with those of Olum (1956, 1958) who, working with Michotte's (1954) *lancement* display, found that the locus of fixation determined whether the two elements were seen to cross or to recoil. This ambiguity in phenomenal identity of the elements at the moment of meeting is familiar to those who have worked with human tracking performance on radar displays. Confusions in phenomenal identity are inevitable unless steps are taken to make the elements discriminable on some dimension such as shape, brightness, size, or relative velocity. Should polar projection of such displays make reproduction more accurate, it would probably be for this same reason—that elements at the moment of crossing were of different size, so that the possibility of a confusion in phenomenal identity was decreased.

Considering the meager data upon which *S*s in these experiments had to base their judgments, it is not altogether surprising that accuracy increased with exposure time. With exposure times of 60 sec. two of the *S*s made almost perfect scores. Direct binocular regard of static arrangements gave higher accuracy with an exposure time half as long. It would appear that the use of the kinetic depth effect is not a particularly rapid way of conveying information about the arrangement of objects in depth.

The increase in accuracy with longer

exposure cannot be attributed to the fact that *Ss* saw more revolutions of the pattern before responding. Were this the case, accuracy of performance should have increased with speed of rotation. It appears that *Ss* simply require considerable time in order to make decisions as to the locations of the elements and to check them by looking at their positions at a subsequent time. The very fact that an accurate percept is so slowly built up in this situation may make it of interest to those such as Smith (1957) who have been searching for ways to study visual recognition as a process extended in time.

Summary

Watching the moving shadows cast by five elements mounted in a 4 × 5 rectangular matrix on a revolving turntable, *Ss* were asked to reproduce the patterns under various display conditions. Only after the introduction of reference markers at the head and foot of the matrix were *Ss* able to reproduce the patterns with an accuracy which exceeded chance expectancy. Using these reference markers, it was found that accuracy of reproduction was increased by increasing the exposure time and by making the elements discriminably different in shape. Size and spacing of the elements, speed of rotation, eccentricity and direction of movement, and giving *S* control over the movement had little or no effect on accuracy of reproduction. There were marked and consistent individual differences throughout the course of the experiments.

References

Benussi, V. Zur experimentellen Grundlegung hypnosuggestiver Methoden psychescher Analyse. *Psychol. Forsch.*, 1927, **9**, 197–274.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Gibson, J. J. Optical motions and transformations as stimuli for visual perception. *Psychol. Rev.*, 1957, **64**, 288–295.

Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

Green, B. F., Jr. Some conditions for the occurrence of the kinetic depth effect. *Amer. Psychologist*, 1958, **13**, 406. (Abstract)

Metzger, W. Tiefenerscheinungen in optischen Bewegungsfeldern. *Psychol. Forsch.*, 1934, **20**, 195–260.

Michotte, A. *La perception de la causalité*. Louvain: Publ. Univer. Louvain, 1954.

Musatti, C. L. Sui fenomeni stereocinetici. *Arch. Ital. Psicol.*, 1924, **3**, 105–120.

Olum, V. Developmental differences in the perception of causality. *Amer. J. Psychol.*, 1956, **69**, 417–423.

Olum, V. Developmental differences in the perception of causality under conditions of specific instructions. *Vita Humana*, 1958, **1**, 191–203.

Philip, B. R., & Fisichelli, V. R. Effect of speed of rotation and complexity of pattern on the reversals of apparent movement in Lissajou figures. *Amer. J. Psychol.*, 1945, **58**, 530–539.

Smith, G. Visual perception: An event over time. *Psychol. Rev.*, 1957, **64**, 306–313.

Wallach, H., & O'Connell, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205–217.

Wallach, H., O'Connell, D. N., & Neisser, U. The memory effect of visual perception of three-dimensional figures. *J. exp. Psychol.*, 1953, **45**, 360–368.

Wallach, H., Weisz, A., & Adams, P. A. Circles and derived figures in rotation. *Amer. J. Psychol.*, 1956, **69**, 48–59.

(Received August 1, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 12–17

POSITIVE AND NEGATIVE KNOWLEDGE OF RESULTS ON A PRESSEY-TYPE PUNCHBOARD

WALTER KAESS AND DAVID ZEAMAN

University of Connecticut

A list of dependable principles of human learning is likely to include some form of Judd's conclusion of 1905 that performance improves most rapidly when *S* is given knowledge of results. The validity of the principle has been demonstrated on tasks as diverse as perception, judgment, sensory discrimination, motor skills, and concept formation. Ammons (1956) found the literature sufficiently coherent to propose 11 tentative generalizations or subprinciples. Wolfle, when summarizing the applications of learning principles states, "laboratory studies are unequivocal in emphasizing the importance of giving a subject as specific and as immediate information as possible concerning the outcome of his efforts" (1951, p. 1267). Both Ammons and Wolfle wisely postpone a discussion of what classes of information constitute effective knowledge until more data is available.

The work of Pressey, who for some 30 years has used several devices which provide human *S*s with immediate knowledge of results, has recently attracted the attention of psychologists interested in "teaching machines." The work with multiple-choice questions keyed to a punchboard which provides knowledge of results is most extensively described in Pressey (1950); and the related work of Angell and Troyer (1948), Jensen (1949), and Little (1934) has been evaluated by Porter (1957) and Skinner (1958).

Porter and Skinner acknowledge the pioneering nature of Pressey's device, but they find fundamental weaknesses in it as a "teaching machine." The assumed weaknesses are based not upon research on the punchboard but rather upon assumptions which the authors have made concerning the most effective method of giving *S* knowledge of results. Both writers emphasize the importance of *S* making a physical response, one which will provide behavioral feedback. Porter, who accepts the punchboard as a "teaching machine" is the more emphatic. Any device which presents *S* with a stimulus but which does not require observable motor action is merely a teaching aid since, "it must be supplemented by some means, usually a teacher in order to be effective" (Porter, 1957, p. 127). Both writers also assume that knowledge of results is most effective when *S*'s behavior is always or nearly always correct. This assumption applies to the punchboard in two ways: the value of alternatives and programming. Porter mentions the "conditioning of incorrect answers" and Skinner (1958, p. 970) states, "effective multiple-choice material must contain plausible wrong responses, which are out of place in the delicate process of 'shaping' behavior because they strengthen unwanted forms." Skinner has another reason for classifying the punchboard as primarily a self-testing device rather than a "teaching machine"—it violates his programming dictum which is, "each step must be so small that it always can be taken, yet in taking it the student moves somewhat closer to fully competent behavior"

(Skinner 1958, p. 970). The punchboard had been used with difficult material which *S* learned by the elimination of incorrect choices.

This paper is primarily concerned with two aspects of the punchboard: the types of information available to *S*, and the importance of behavioral feedback. A multiple-choice item keyed to a punchboard can provide two types of knowledge: positive, when *S*'s punch perforates the paper signifying "this is the correct answer," and negative, when the punch does not break the paper and signals "this is not the correct answer." The experimental strategy is to hold constant the amount of positive information *S*s receive and to systematically vary the amount of negative information. The importance of behavioral feedback is studied by making its presence or absence an independent variable. The data can be used to test some of the assumptions made by Porter and Skinner as to the nature of effective knowledge of results.

METHOD

Experiment I.—Each *S* received a test booklet containing 150 multiple-choice items and an IBM answer sheet stapled to a cardboard template which provided immediate knowledge of results. The test items concerned the definition of psychological terms. One set of 30 items was used throughout the test. Trial 1 consisted of Items 1–30; Trial 2, Items 31–60, the original items in a new sequence; Trial 3, Items 61–90, a different sequence of the same items, etc. The sequence of items for each trial was determined by a table of random numbers, as was the order of choices within each item. Thus serial-position effects for the items and position effects for the placement of the correct answer within an item were controlled.

The independent variable, amount of negative information, was manipulated by varying the number of item-choices available to *S*s on Trial 1. Condition 5 offered 30 5-choice items; Cond. 4, 30 4-choice items; Cond. 3, 3-choice items; and so on for Cond. 2 and 1. The elimination of a distractor for a given condition was made at random from the next highest condition number. See Table 1 for the form used. This procedure reduced the possibility of systematically eliminating attractive or unattractive distractors. Trials 2–5 used only 5-choice items which were the same for all conditions.

The IBM answer sheet (1TS 1000B 108) was converted into a punchboard by means of a cardboard template containing 150 ¼-in. holes which were positioned so as to be directly beneath the correct item-choice. A blank sheet of paper was inserted between the answer sheet and the template to reduce tactile and visual cues. This assembly was then stapled to a solid sheet of cardboard.

The experimental conditions were determined by the test booklet *S* received. The *S*s, students enrolled in the introductory psychology course, were tested in groups of about 30. All conditions were used in each student group, the test booklets being distributed in a predetermined random fashion. The random assignment of conditions within groups equated *S*s in the different conditions for initial knowledge of psychology. The *N*s for each condition were: 5NC, 88; 5, 75; 4, 64; 3, 62; 2, 73; 1, 73.

The control condition, 5NC, no correction, on Trial 1 used a modified version of the punchboard. The template was solid beneath the items of Trial 1. The *S*s marked their

TABLE 1

FORMS OF CHOICE PRESENTATION OF ONE ITEM FOR SIX CONDITIONS
(Item: A method of studying the interrelationships within a social group)

Cond.	Choice Presentation				
5 & 5NC	1 psychograph;	2 interaction meter;	3 paired comparisons;	4 forced choice;	5 sociometry
4	1 psychograph;	2 blank;	3 paired comparisons;	4 forced choice;	5 sociometry
3	1 psychograph;	2 blank;	3 paired comparisons;	4 blank;	5 sociometry
2	1 psychograph;	2 blank;	3 blank;	4 blank;	5 sociometry
1	1 blank;	2 blank;	3 blank;	4 blank;	5 sociometry

choices in the conventional manner by a vertical line between the choice markers and received no knowledge of results. Trials 2–5 used the holes in the template as a punchboard, and thus gave knowledge of results.

Experiment II.—An effort was made to evaluate the importance of behavioral feedback, the reinforcement value of punching on a punchboard. Test booklets for the first trial of Cond. 1 and 5 were modified by underlining with a red pencil the correct answer. The *Ss* were instructed to read carefully all available choices and to note the choice underlined since it was the correct answer. The *Ss* made no use of the punchboard until they began Trial 2. These modifications constitute Cond. 1A and 5A. Condition 5B was established by using the form of test booklet used for Cond. 5A, but *Ss* were instructed to punch the correct choice on the answer sheet. All *Ss* worked on the same material on Trials 2–5. The *N*s for each condition were: 1A, 49; 5A, 30; 5B, 40.

Results

Presence or absence of knowledge of results.—The *Ss* in Cond. 5 and 5NC took the same tests on Trials 1 and 2. On Trial 1 the punchboard provided *Ss* in Cond. 5 with knowledge of results, while *Ss* in Cond. 5NC took a multiple-choice test in the usual manner. As can be seen from Fig. 1, there is little overlap in the distribution of errors made by the two groups on Trial 2. The t of 27.86 indicates the magnitude of the effect of the modified punchboard with vocabulary items.

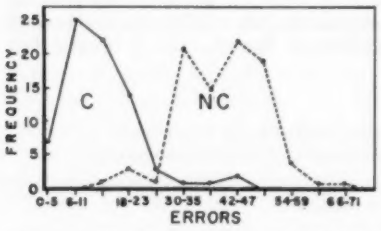

FIG. 1. Distribution of errors on Trial 2. On Trial 1, *Ss* in Cond. 5NC, no correction, took multiple-choice test without knowledge of results while *Ss* in Cond. 5 received knowledge of results from 5-choice punchboard.

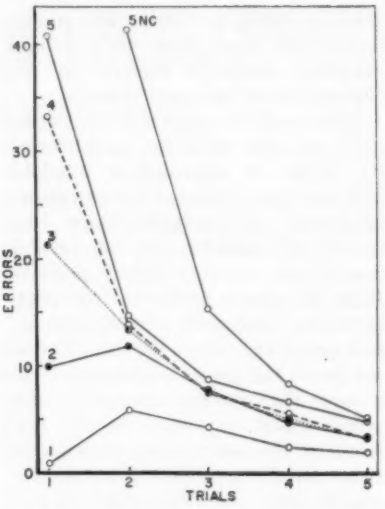

FIG. 2. Mean errors made on punchboard by groups having 1 to 5 item choices available on Trial 1. Items for Trials 2–5 had 5 choices.

Negative knowledge.—On the punchboard *S* receives negative knowledge when he punches an item-choice and is informed "this is not a correct answer." The mean number of errors on Trial 1 for each condition in Fig. 2 can be interpreted as the number of times the "average" *S* in the group received negative knowledge. All groups received the same amount of positive knowledge since they punched each item until there was a perforation signaling the correct answer. Errors on Trial 2 can be taken as the dependent variable. It appears from Fig. 2 that the more negative knowledge a group receives the more errors it makes. A simple analysis of variance of Trial 2 errors, control group excluded, gave an F of 13.36 ($df = 4, \infty$; $P < .001$).

Negative knowledge could produce a relative increase in errors on subsequent trials in at least two ways: (*a*) negative information makes it

less likely that *S* will mark the correct answer again and *S* distributes his picks randomly among the remaining available choices; (*b*) negative information makes it more likely that *S* will repeat the original error. Condition 2, where *S* is presented with but two choices, one being correct, is the easiest place to evaluate the two explanations. For each item of Trial 1, *S*s can be classified into a "rights" group which chooses the correct answer and receive positive information, and a "wrongs" group which selects the choice we will term the prime distracter and receives negative information (and on the second try receives positive information). The "rights" and "wrongs" groups can be further classified by their performance on the same item on Trial 2. The behavior of the "wrongs" group is pertinent. The essentials of 30 separate item analyses can be summarized by using the mean percentages of the Trial 2 response categories: correct 53%; prime distractor alone 27%; prime distractor and other errors 15%; errors but not prime distractor 5%. Since the prime distractor for each item was selected by using a table of random numbers there is no reason to believe that these 30 distractors are more attractive choices than the population of distractors from which they were selected. If we assume that 53% of the "wrongs" group learned the correct answer on Trial 1 we would expect about 12% of the errors to be assigned to each of the *four* distractors on Trial 2. The conclusion is that negative information increases the probability that *S* will repeat his error.

The item analysis suggested that item difficulty may be related to the effect of negative information. Negative information seemed to have little effect on easy items which were missed but a strong effect on difficult items. This can be expressed by using the number of *S*s in the "rights" group as an index of the difficulty of an item. A second index, one of the effectiveness of negative information, is based upon the percentage of the "wrongs" group which answers an item without error on Trial 2. The product-moment correlation between the two indexes is +.62. This tentative conclusion can not be adequately tested with our data. The test items were purposely made difficult, but fortunately in some cases we underestimated the student's psychological vocabulary.

Repeated trials.—Previous studies of learning on the punchboard seem to have used only a single time through the test material. Repeated trials could result in the rapid reduction of errors or a slow reduction of errors to an asymptote considerably above perfect performance. The shape of the curve would likely depend upon the relative strengths of positive and negative knowledge. The error curves of Fig. 2 bear on this matter. The curve for Cond. 5 shows a rapid reduction in errors for the early trials but seems to be flattening out in the later trials. After completing four trials with 5-choice items the *S*s in Cond. 5 are making about the same number of errors as *S*s in Cond. 1 make after one trial. The strong resistance of the effects of negative knowledge to correction with a 5-choice punchboard is shown not only in Fig. 2 but also in Fig. 3, where the errors on Trials 2–5 are the result of varying the amount of negative information on Trial 1. It seems reasonable to conclude that the 5-choice punchboard rapidly reduces errors on beginning trials but soon reaches a point of diminishing returns,

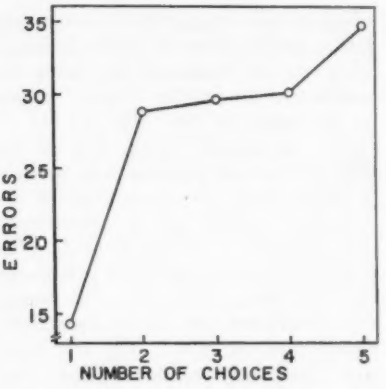

FIG. 3. Mean total errors on Trials 2–5 as a function of the number of item choices available on Trial 1.

Part of this effect is due to the negative information received.

Behavioral feedback.—The relative importance of overt vs. covert behavior remains one of the fundamental questions of learning. Experiment II was designed to evaluate the importance of the punchboard. The results of Exp. II and Cond. 1 of Exp. I, given in Table 2, were analyzed by a double classification analysis of variance, using adjustment for unequal numbers. The main effects between Punch-Observe and between 1–5 Choices and their interaction did not remotely approach conventional levels of statistical significance.

TABLE 2

ERRORS MADE ON TRIAL 2 BY FOUR GROUPS WHO OBSERVED OR OBSERVED-AND-PUNCHED CORRECT ANSWER ON TRIAL 1 WHEN PRESENTED WITH 1- OR 5 ITEM CHOICES

Choices	Observe-Punch				Observe Only			
	Cond.	*N*	Mean	*SD*	Cond.	*N*	Mean	*SD*
1	1[a]	73	5.89	5.99	1A	49	4.51	4.75
5	5B	40	6.10	6.77	5A	30	6.10	6.65

[a] From Experiment I.

The similarity of the mean scores of the experimental conditions, Table 2, suggests that *S*s received only positive knowledge. The *S*s seem to neither profit from nor are they hindered by the number of choices. The presence or absence of overt behavior does not seem to influence the learning. If this information is combined with the demonstrated effect of punching an incorrect choice found in Exp. I, we reach an interesting hypothesis: If an incorrect answer, as well as a correct answer, is read under instructions to note especially the underlined answer, the incorrect answer does not "stick." But if an incorrect answer is punched, in the absence of underlining of the correct answer, the incorrect answer tends to "stick" as a correct answer, but even so may be overcome on later trials. Alternative explanations may be found but there are data which further support the conclusions used in the chain of reasoning; namely, the essential equality of performance of the four groups used and the fact that the curves for each of the three conditions of this experiment are for Trials 2–5 almost identical with that of Cond. 1 of Fig. 2.

DISCUSSION

The two kinds of knowledge of results, positive and negative, when presented through tactile and kinesthetic cues on the punchboard, differentially effect performance. Knowledge that an item choice is correct reduces errors; information that an item-choice is not a correct answer leads to a relative increase in errors beyond that observed without such negative information. The assumption of Porter and Skinner that incorrect items interfere with the acquisition of correct responses is supported by the results of Exp. I. However, their analysis is not supported by either of the two findings of Exp. II. Negative

information does not impair performance when it is presented for *S* merely to read, nor is the reinforcement of overt correct responses more effective than the reinforcement of covert behavior.

An interpretation of the interfering effects of negative information requires identification of the source of reinforcement for the observed tendency to repeat erroneous choices. Two possible hypotheses that may serve as a guide for future research are as follows: (*a*) The temporal contiguity of question stimuli and erroneous responses provides a sufficient condition for their connection. One difficulty with this interpretation is the fact that the punchboard requires the last or postreme response to every question to be correct. (*b*) A reinforcement interpretation may appeal to the reward-value of the punch feedback following a correct response. This may provide a delayed reinforcement for incorrect responses preceding the correct one. Since the effects of delayed reinforcement vary with the length of delay, a testable consequence of this latter view, at least, is the prediction that temporal spacing of questions and answers will be a parameter of the main effects of the present experiment.

Further experimental analysis of the efficiency of the punchboard as a training device may yield additional results of both theoretical and applied interest.

Summary

A modified Pressey-type punchboard was used to study the influence of positive and negative knowledge of results and the importance of punch feedback. The rate of multiple-choice learning of definitions of psychological terms was found to vary inversely with the number of available wrong answers when overt error responses (punches) were made, but not when wrong answers were merely read. No corresponding differential effect of punch feedback on correct answers was observed. Reading these was as effective in producing learning as reading and punching.

References

Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *J. gen. Psychol.*, 1956, **54**, 279–299.

Angell, G. W., & Troyer, M. E. A new self-scoring device for improving instruction. *Sch. Soc.*, 1948, **67**, 84–85.

Jensen, B. T. An independent-study laboratory using self-scoring tests. *J. educ. Res.*, 1949, **43**, 134–137.

Little, J. K. Results of use of machines for testing and for drill, upon learning in educational psychology. *J. exp. Educ.*, 1934, **3**, 45–49.

Porter, D. A critical review of a portion of the literature on teaching devices. *Harv. educ. Rev.*, 1957, **27**, 126–147.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, **29**, 417–447.

Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

Wolfle, D. Training. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1267–1270.

(Received July 2, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 18–22

REMEMBERING THE PRESENT STATES OF A NUMBER OF VARIABLES[1]

DOUWE B. YNTEMA AND GAYLE E. MUESER

Lincoln Laboratory, Massachusetts Institute of Technology

Card players, air-traffic controllers, and people going about their ordinary business demonstrate an ability to keep track of a number of things at once. It seems plausible to study this sort of memory in the following way: The *S* is read a series of messages. Each tells him that one of a number of variables is and until further notice will remain in some particular state. At random intervals the series of messages is interrupted and he is asked what the present state of one of the variables is.[2]

Two questions are investigated here: How does performance depend on the number of variables? And, is keeping track of several variables whose states are all drawn from the same set of possibilities different from keeping track of an equal number of variables that have their own, mutually exclusive sets of possible states?

METHOD

Subjects.—The *S*s were eight male Air Force lieutenants assigned by a superior officer. Their ages were 22 to 27 yr., they had had no experience with experiments of this type, and they had all been trained as Intercept Controllers.

Apparatus.—The *S* sat before a large sheet of white plastic on which were fastened a number of opaque doors, one for each variable. He wrote on the plastic with a grease-pencil and erased with a rag.

Lists of stimuli.—It is convenient to think of each variable as an attribute of an abstract, nonsensical object. The objects are identified by letters: D, H, K, L, N, W, X, and Z. The eight attributes and their states are shown in Fig. 1. Thus the "shape of D" would be a typical variable and "circle" would be one of its states.

The list of stimuli for an experimental run had three parts, which will be called the prestarters, the starters, and the body. The body consisted of 128 items, each chosen independently of the others by a random process equivalent to the following: First, it is decided whether the item will be a message or a question. The probability of a question is one-half. Next, the variable to which the message or question will apply is chosen, all variables being equally likely. And finally, if the item is to be a message, one of the four possible states of the chosen variable is selected, all states being equally likely. The starters consisted of one message about each variable. They were chosen by selecting a random permutation of the variables and picking at random one of the four states of each. The prestarters also consisted of one message about each variable. They were chosen by taking the variables in the order in which they were assigned to the doors and picking at random one of the four states of each. The lists were generated by an IBM Type 704 computer programmed to follow the rules just described.

Procedure.—Before a run each door was labeled with the name of a variable. A row of doors represented an object and a column an attribute, or vice versa. The order in which attributes were assigned to doors was, reading left to right or top to bottom, the order obtained by deleting from Fig. 1 any attributes that were not to be used. The order of the objects was alphabetical. The *S* practiced writing the states of all attributes that were to be used. When *E* was satisfied that the whole set of symbols could be recalled without hesitation, he read the prestarters and *S* recorded the appropriate

[1] The work reported in this paper was performed by Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the United States Army, Navy, and Air Force.

[2] A somewhat similar experimental situation has been developed independently at the University of Virginia (Lloyd, Reid, & Feallock, 1960).

FIG. 1. The so-called attributes and the four states of each.

symbol behind each door. The *S* was told he need not try to remember the prestarters.

After those preparations the task began. The *E* read the starters and the items in the body of the list one by one. Messages were spoken in the form, "D circle," and questions in the form, "Shape of D?" When *E* read a message *S* repeated it aloud and opened the door to which it referred. If the symbol behind the door agreed with the message, he simply closed the door again. If the symbol and the message did not agree, he erased the old symbol, substituted the new one, and closed the door. In either case *E* read the next item as soon as the door was closed. When *E* read a question *S* attempted to say what was behind the door to which it referred, guessing if necessary. As soon as he responded, *E* read the next item.

The *S*'s answers were not corrected and he was not allowed to open the doors except to post messages. He was encouraged to work rapidly and not to worry about the states of the other variables while posting a message or answering a question. With these instructions the average length of a run was about 8 min. Runs were done two per session with a rest of about 25 min. between sessions. There were a practice session and two experimental sessions on the first day, and three experimental sessions on the day following. All *S*s used the same practice lists. The first was a half-length list with four variables: direction and animal of K and X. The second was a full-length list with nine variables: shape, mark, and weather of H, N, and Z.

Conditions and design.—There were 10 experimental conditions: 2, 3, 4, 6, and 8 objects with the same single attribute; and 2, 3, 4, 6, and 8 attributes of one object. Each *S* made one run under each condition. The order of conditions was balanced. Attributes were assigned to runs in such a way that in any given condition every attribute was used in as many runs as every other attribute, and the letters designating objects were assigned in the same way. With four *S*s, chosen at random, a row of doors always represented an object and a column an attribute. With the other four this arrangement was reversed.

Since it was anticipated that some of the randomly generated lists of stimuli would be more difficult than others, the list that an *S* was presented in a many-attribute condition and the list that he was presented in the one-attribute condition with the same number of variables were, unknown to him, made as nearly alike as possible. In effect, the computer generated 40 independent lists of stimuli for the 40 many-attribute runs and then translated each into a one-attribute list with the same number of variables. The translation was done by putting the variables into one-to-one correspondence with each other and then putting into one-to-one correspondence the messages about corresponding variables. Questions were translated into questions about corresponding variables, and messages into the corresponding messages about corresponding variables. The translated item occupied in the new list the same position that the original item occupied in the old one.

RESULTS

Figure 2 shows the principal results. As might be expected, the fraction of the questions answered correctly decreases as the number of variables whose states are to be remembered increases. What is more interesting, the fraction correct is lower when the variables are all one attribute of different objects than when they are different attributes of the same object. The difference between the two curves is not significant at two variables; however, two-tailed *t* tests in which each *S* was matched with

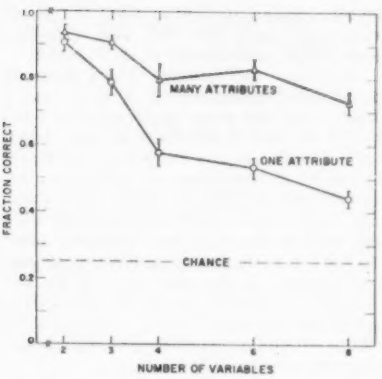

FIG. 2. Relation between fraction of the questions answered correctly and number of variables whose states were to be remembered. In the curve labeled "one attribute," all variables have the same set of possible states; in the curve labeled "many attributes," each has its own exclusive set of states. The points are means of the scores of the eight *S*s. The bars extend $\pm 1\ \sigma_m$.

himself showed significant differences at 3, 4, 6, and 8 variables (*P*s less than .001, .01, .001, and .001, respectively), and two-tailed sign tests gave the same result (all four *P*s less than .01).

Figure 3 shows how the probability of a correct response is affected by the number of messages and questions that happen to intervene between the question under consideration and the message that determines the correct answer. The figure makes apparent one reason why the fraction correct was found to decrease as the number of variables increases: when there are many variables the number of intervening items is frequently large, and, as the figure shows, the probability of a correct response drops as the number of intervening items increases.

The four points on the curve for a given experimental condition in Fig. 3 were determined in the following way: The questions from all runs done under the given condition were tallied by number of intervening items. With the tallies from all *S*s pooled, number of intervening items was divided into four class intervals such that each would contain about 128 questions. The data for the various *S*s were then separated again and two statistics were computed in each interval for each of the eight *S*s: first, the mean number of intervening items for the questions that fell in the interval; and second, the fraction of these questions answered correctly. The mean of the eight means is the abscissa of one of the points plotted on Fig. 3, and the mean of the eight

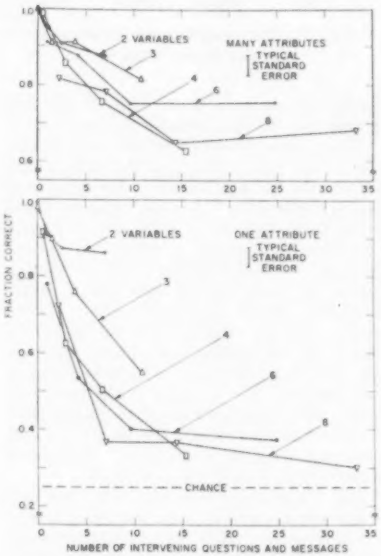

FIG. 3. Fraction of the questions answered correctly, shown as a function of number of messages and questions intervening between the question under consideration and the message that determined the correct answer. The parameter is the number of variables whose states were to be remembered. The length of the vertical bar is one standard error of the typical point below .9 on the panel on which it appears.

fractions correct is the ordinate. The standard error of estimate of the ordinate was determined from the same eight fractions. Since ordinates close to 1.0 tended, almost of necessity, to have small standard errors, the median of the standard errors of ordinates that fell below .9 is shown as typical.

To learn what should be controlled in an experiment like this, an attempt was made to identify the sources of the error variance represented by the vertical bars in Fig. 2. With these Intercept Controllers the variance due to *S*s was not significant and accounted for only a trivial fraction of the error; the order of conditions also accounted for only a trivial fraction of the error; and the attributes did not differ significantly in difficulty. Only one source of error was identified: interaction between *S*s and number of variables was significant ($P < .05$) and accounted for roughly one-third of the error variance. This interaction may well arise from differences between lists of stimuli: as has been explained, each randomly generated pair of similar lists was used by only one *S* and at only one level of number of variables.

Discussion

It is surprising that keeping track of several attributes of one object should be so different from keeping track of one attribute of several objects. The two tasks are formally the same; in fact, the lists of stimuli were identical, item for item, except for the names of the variables and of their states. There seem to be two directions in which the explanation might be sought.

The first approach involves Miller's (1956) concept of the *chunk*—an item, like a word or a digit, that is remembered as a unit. Suppose that *S* is dealing with several attributes of one object and is asked, "Shape of D?" If he can remember that the present state of one of the variables is "circle," then he has all the information that he needs to answer correctly. On the other hand, remembering "circle" is not enough when he is dealing with several objects that all have the one attribute, shape. Now he needs the whole phrase, "D circle." Thus it could be argued that in the one-attribute case he is forced to code into two chunks, analogous to "D" and "circle," the same information that could be coded into a single chunk, "circle," in the many-attribute case. Miller (1956) points out that the number of chunks in the memory span is approximately independent of the number of bits of information per chunk, and he suggests that in general a given amount of information is easier to handle when coded into a few chunks than when coded into many.

The second approach is to ask how the items that are to be remembered might interfere with each other. In the many-attribute case the present states of the variables are all from different rows of Fig. 1, and thus form a rather heterogeneous set of items. They are "isolated" from each other in the way that von Restorff (1933) found helpful in a variety of learning tasks. In the one-attribute case the present states are all from the same row of Fig. 1 and are, by contrast, similar to each other—sometimes identical. Thus it could be argued that they form a homogeneous set of the sort that von Restorff found is more difficult to learn. Or consider the A-B, A-Br paradigm for associative interference in paired-associate learning. If the response R_1 has been paired with Stimulus S_1 and R_2 has been paired with S_2, then (provided that the responses are familiar words) there is considerable negative transfer to learning R_2 as a response to S_1 (Besch & Reynolds, 1958; Porter & Duncan, 1953; Twedt & Underwood, 1959). Analogous changes in the pairing of states and variables occur frequently in the course of a one-attribute run: the present state of one variable is often the same as a past state of another

variable. But in the many-attribute case each variable has its own set of states, and this sort of re-pairing cannot occur.

Summary

The task was to remember the present states of 2, 3, 4, 6, or 8 variables. The *S* was read a series of messages, each saying what state one of the variables had assumed. He recorded the messages on a board on which the space for each variable was covered by a door that was open only while messages about that variable were being posted. The series of messages was interrupted at random by questions, each asking what was behind one of the doors. In one case (called the one-attribute case), all variables had the same set of four possible states; in the other (called the many-attribute case), each variable had its own exclusive set of four states.

The fraction of the questions answered correctly decreased as the number of variables increased and, if there were more than two variables, was lower when they all had the same set of four possible states. The probability that a particular question would be answered correctly depended on the number of questions and messages intervening between the question under consideration and the message that determined the correct answer.

References

Besch, N. F., & Reynolds, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 554–558.

Lloyd, K. E., Reid, L. S., & Feallock, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, in press.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

Twedt, H. M., & Underwood, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

von Restorff, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychol. Forsch.*, 1933, **18**, 299–342.

(Received July 6, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 23–29

STIMULUS–RESPONSE GENERALIZATION WITH DISCRETE RESPONSE CHOICES[1]

GUSTAV LEVINE

Creedmoor Institute for Psychobiologic Studies

If *S*s are taught two mutually exclusive responses to two different stimuli from a single stimulus dimension, the presentation of stimuli intervening between the two conditioned stimuli can create conflict. If the mutually exclusive responses are actually responding versus not responding, a generalization gradient is generally found. That is, the single response can vary on a continuum of frequency or intensity. Underwood (1949, p. 247) has indicated that a compromise response frequently occurs in conflict situations such as the Hovland and Sears "conflict board," although there is no provision for the compromise response in the instructions. The compromise response in conflict experiments could be seen as similar to the less intense or less frequent response in those generalization experiments in which the response continuum lies between responding and not responding. If the generalization response is a compromise response, this could only become clear in an experiment in which there is a choice of more than two responses which have some continuum qualities in common. In a situation where there are a limited number of responses to choose from, the responses can be differentiated along a continuum on the basis of some stimulus attributes inherent in the way in which the responses are made available. For example, a group of tasks may be located at several points in space, or presented sequentially in time, and each response can then be seen as related to the others in that it can be compared with them on the same dimension (more to the left, or later, etc.).

In the present experiment, a group of nonsense syllables, representing response choices, were each made available to *S*s on cards which were spread out on a horizontal plane, presenting the possibility of identifying the responses as members of a spatial continuum, from left to right or vice versa. The two end responses of the response continuum were then conditioned to the two end stimuli of a stimulus continuum, which presented the possibility of examining the effects of generalization on discrete response choice, when the intervening stimuli were later presented. The compromise responses were made available along with the conditioned (end) responses during the test of generalization. In this way the end responses and the intervening compromise responses could be compared for frequency of occurrence in the presence of the intervening stimuli. The major purpose of this experiment is to measure any possible change effected in nonconditioned intervening responses (in the presence of nonconditioned intervening stimuli), when conditioning has taken place at the ends of the stimulus and response continua. The experiment was done

[1] The author wishes to express his gratitude to the Director of the Institute, John R. Whittier, and to Harry A. LaBurt, Senior Director of Creedmoor State Hospital, for providing the atmosphere and the support which made this experiment possible.

twice, with two different stimulus dimensions.

The increase in frequency of an intervening response in the presence of an intervening stimulus may be a function of any new stimulus, not just an intervening stimulus. Therefore, two intervening stimuli have been used, along with two intervening response choices. Each intervening stimulus is closer to one of the end stimuli, and each intervening response is closer to one of the end responses, making for a set of four ordered stimuli and four ordered responses, which can be placed in a one to one correspondence based on ranks. A meaningful increase in intervening responses to intervening stimuli would therefore require that the increase be specific to "appropriate" stimulus-response pairs (the word appropriate referring to stimuli and responses that would correspond to each other in a ranking of both the set of stimuli and the set of responses).

It has been hypothesized that when the end responses have been conditioned to the end stimuli, there is an increase in the frequency of nonconditioned intervening responses to the appropriate nonconditioned intervening stimuli.

Method

Experiment A

Subjects and procedure.—The *S*s were 20 male college students. They were told that the experiment was a study of factors in learning. They were also informed that to help them maintain interest in the task, which was set up as a guessing game, they would receive a penny for every correct guess, up to a dollar.

Preconditioning series.—The *S*s were handed four 3 × 5 in. white cards, placed together in random order, with a different nonsense syllable typed on each card. They were told to spread the four cards out in front of them, leaving room below for working on the desk. The *E* kept emphasizing the need for room below until the cards were placed in a horizontal order. When the cards had been properly arranged, a new set of cards with small tabs at their centers, used as stimuli, were presented to *S*.

The specific task for *S*, during the initial phase of the experiment, was the choice of one nonsense syllable, from among the four nonsense syllables typed on the cards. The *S*s were asked to choose a nonsense syllable, pronouncing the syllable, each time that they were presented with one of the cards which contained a small tab pasted on its center. The cards with the tabs were taken, one at a time, from a deck of 40 such cards, containing four different sizes of tabs. The *S*s were told that beneath each tab was a nonsense syllable (one of the four syllables placed on the four response cards in front of them). The fact that there were differences in the tab sizes on the stimulus cards was not pointed out to *S*. He was merely asked to guess which word was under the tab on each card. Each *S* was informed that at first there would be no indication from *E* as to whether or not each guess was correct, but that later he would be told when he was correct, and we would see how this affected his guesses. This initial phase in which *S* guessed without reward permitted the establishment of each *S*'s operant level preferences.

Conditioning series.—After the 40 cards with four different tab sizes had been presented to *S*, a new set of 40 cards was brought forth for the purposes of conditioning. This second deck had only the largest and smallest tabs on its cards, 20 of each size. The *E* removed the two center response choice cards, from the four response cards in front of *S*, leaving only the two end cards. These end responses were pushed together by *E* while he explained to *S* that with the new deck only the two remaining nonsense syllables were relevant. The *S* was told that he would be informed that he was correct each time that he guessed correctly which of the two syllables was beneath the tab of the presented card, and reminded that he was getting a penny for each point (correct guess). He was given a score card which was numbered from 1 to 100, and told that he was to cross out a number each time that *E* said "point," giving *S* his own cumulative record of points. The conditioning cards were then presented. For the first 10 *S*s the *E* responded with the word "point" each time *S* responded with the syllable on the left in the presence of the small tab, and also when *S* responded with the syllable on the right in the presence of the large tab. For the next 10 *S*s the *E* reversed the conditioned

relationship, left syllable with large tab, right syllable with the small tab.

The conditioning procedure was continued until at least all 40 cards were presented, and stopped at that point if the last 20 responses were correct. If not, the deck (or half of the deck) was again presented until either the first 20, or the last 20 cards were responded to without error. If this did not occur after four presentations of the complete deck, *S* was dropped from the experiment. Three *Ss* had to be replaced because they had not met the criterion in these 160 trials.

Postconditioning series.—After the conditioning series, the preconditioning deck and response choices were reintroduced. The *S*'s score card was removed, and the previously removed center nonsense syllables were put back in their original places on the desk by *E*, so that there were now the original four responses available as choices. The *S* was told that he would be working with the first deck again, in order to see what he had learned from the experience. He was told that although he would not be informed now as to whether his guesses were right or not, *E* would continue to keep score. The preconditioning deck with the four different tab sizes was then presented.

Materials.—The tabs on the stimulus cards were $\frac{1}{4}$ in. wide, and were either $\frac{3}{8}$, $\frac{5}{8}$, $\frac{7}{8}$, $1\frac{1}{4}$ in. in length. The tabs were cut from the same material as the cards. The preconditioning deck contained the four tab sizes in random order. The conditioning deck, which contained only the smallest and largest tabs, were placed in random order with the restriction that the first 20 cards contained 10 of each tab size.

The response choices consisted of four 3 × 5 in. cards, white and unlined, each containing a nonsense syllable of 5% association value as determined by Hull (1933). The syllables were ZUD, NAF, PIJ, and MEV, typed in capital letters, and were each placed in the center of a card.

Experiment B

Subjects and procedure.—The *Ss* were 20 additional male college students. The procedure and response choices were the same as in Exp. A. However, different stimuli were used. The first 10 *Ss* were conditioned to respond with the nonsense syllable on the left to the lightest tab, and with the nonsense syllable on the right to the darkest tab. The next 10 *Ss*, as in Exp. A, had the stimulus-response relationships reversed.

Materials.—The preconditioning deck (also used as the postconditioning deck) consisted of 40 3 × 5 in. white unlined cards. In the center of each card there was a piece of rectangular cardboard, $\frac{1}{4}$ in. wide and $\frac{1}{2}$ in. long. Ten of the center cardboard tabs were made of the same white material as the 3 × 5 in. cards, 10 were cut from gray color-aid paper,[2] with Color Code No. 1, 10 were cut from Color Code No. 3, and 10 were cut from Color Code No. 8. Color Code No. 8 is almost black, and 1 and 3 are intermediate shades of gray. The four kinds of tabs therefore ranged from white, through two shades of gray, to almost black. The differently shaded stimuli were arranged in the deck in a random order.

The conditioning deck was made of the same materials and in the same way, as the preconditioning deck. However, only the lightest and darkest tabs were in the conditioning deck, 20 of each, placed in a random order, with the restriction that the first 20 cards contained 10 of each tab color.

Results

The statistical test used throughout the treatment of the data was the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956). Significance was computed in terms of two-tailed tests, and probabilities greater than .05 were considered nonsignificant.

The phrase "appropriate responses" has been used below to denote, in the case of the first 10 *Ss* of Exp. A, the response on the card furthest to the left of *S* to the smallest tab, the response second from the left to the second smallest tab, the response second from the right to the second largest tab, and the response furthest to the right to the largest tab. For the next 10 *Ss* the meaning of appropriate responses was reversed, in keeping with the reversed conditioning. For the first 10 *Ss* of Exp. B, the response furthest to the left when evoked by the lightest stimulus, is considered an appropriate response, as is the response second from the

[2] The color-aid paper was obtained from Rosenthal Artist Materials, 840 Broadway, New York City.

left to the second lightest stimulus, etc. For the second 10 *Ss* of Exp. B, the meaning of appropriate response is again reversed.

Table 1 presents the total number of choices of each response to each stimulus, comparing the preconditioning to postconditioning totals, for both Exp. A and B.

The primary question in this experiment is whether *Ss* increased their nonconditioned intervening responses to the appropriate intervening stimuli after the conditioning of the end responses to the end stimuli. In Exp. A, 18 of the 20 *Ss*, and in Exp. B, 19 of the 20 *Ss*, increased their intervening responses to the appropriate stimuli, over their preconditioning levels, supporting the hypothesis. Table 2 summarizes the Wilcoxon significance tests.

The *Ss*, during conditioning, were taught to respond to the end stimuli with the end responses. If *Ss* learned that the end responses in some sense belonged to the end stimuli only, so that during the postconditioning series they saw the choice of responses to the intervening stimuli as being between the two intervening responses only, there would have been a one out of two choice of appropriate responses, whereas during the preconditioning series there had been a choice between four responses to a particular stimulus.

TABLE 1

TOTAL CHOICES OF EACH RESPONSE TO EACH STIMULUS FOR ALL *Ss*

Stimuli	Preconditioning Responses Reading Left to Right				Postconditioning Responses Reading Left to Right			
Exp. A: First 10 *Ss*								
Smallest	*41*[a]	29	9	15	*99*	1	0	0
2nd Smallest	31	*17*	25	30	0	*86*	12	0
2nd Largest	14	21	*44*	34	1	12	*82*	7
Largest	14	33	22	*21*	0	1	6	*93*
Exp. A: Second 10 *Ss*								
Smallest	27	22	14	*23*	0	0	13	*88*
2nd Smallest	19	27	*32*	26	0	22	*72*	9
2nd Largest	30	*30*	33	19	5	*74*	15	3
Largest	*24*	21	21	32	*95*	4	0	0
Exp. B: First 10 *Ss*								
Lightest	*30*	11	31	22	*98*	0	0	5
2nd Lightest	21	*28*	18	31	0	*84*	22	0
2nd Darkest	25	32	*26*	28	0	16	*78*	1
Darkest	24	29	25	*19*	2	0	0	*94*
Exp. B: Second 10 *Ss*								
Lightest	26	19	22	*30*	0	0	1	*100*
2nd Lightest	23	21	*18*	31	1	15	*75*	0
2nd Darkest	17	*30*	35	25	0	*82*	24	0
Darkest	*34*	30	25	14	*99*	3	0	0

[a] All "appropriate" responses are italicized.

TABLE 2

T Scores of Appropriate Responses of Postconditioning Minus Appropriate Responses of Preconditioning

Stimuli	*T* Found	*T* Needed[a]	
		.05	.01
2nd Smallest	0.0	52	38
2nd Largest	4.0	52	38
2nd Lightest	0.0	35	23
2nd Darkest	7.5	52	38

[a] Probability of significance increases with decrease in *T*.

This change in number of responses from which to choose could create a spurious increase in the use of the intervening responses to appropriate stimuli. That is, learning involving the end responses and stimuli would be sufficient to explain the change in intervening responses without any implications of learning of intervening stimulus-response relationships. A ratio, involving changes in the use of the intervening responses only, was therefore constructed. The numerator was the number of the appropriate responses to an intervening stimulus, and the denominator was the total number of intervening responses (both appropriate and not appropriate) to that same stimulus. This fraction was computed for each *S* for each of the intervening stimuli, and the change in this ratio from preconditioning to postconditioning was tested for significance. The ratios of the appropriate responses to both stimuli were increased significantly in both experiments, at better than the .01 levels, indicating that the increase in the use of the appropriate intervening responses involved some learning with reference to the intervening stimulus-response relationships, although only the end stimulus-response relationships had been conditioned. Table 3 summarizes the Wilcoxon significance tests.

TABLE 3

T Scores of Ratio of Postconditioning Responses Minus Ratio of Preconditioning Responses

(Ratio consists of sum of appropriate responses to an intervening stimulus over sum of both intervening responses to that stimulus)

Stimuli	*T* Found	*T* Needed[a]	
		.05	.01
2nd Smallest	8.5	40	28
2nd Largest	15.5	52	38
2nd Lightest	11.0	35	23
2nd Darkest	25.5	46	32

[a] Probability of significance increases with decrease in *T*.

It is generally expected that conditioning a response to one stimulus also increases the occurrence of this response to neighboring stimulus values. However, this assumption is generally not tested in the presence of alternate response choices. A test was made of the increase of each conditioned (end) response to the intervening stimulus nearest to the end stimulus to which the response had been conditioned. The results indicate that not only was there no increase of the conditioned responses

TABLE 4

T Scores of Decrease of Conditioned Responses to Nearest Nonconditioned Stimuli

Stimuli Nearest CS	*T* Found	*T* Needed[a]	
		.05	.01
2nd Smallest	0.0	14	7
2nd Largest	18.5	21	13
2nd Lightest	0.0	35	23
2nd Darkest	0.0	35	23

[a] Probability of significance increases with decrease in *T*.

to the nearest nonconditioned stimuli, but that there was a significant decrease. Table 4 summarizes the Wilcoxon significance tests.

Since indirect conditioning appears to have taken place in this experiment, the nonconditioned responses increasing to the nonconditioned stimuli, how does this compare to the increases occurring with the directly conditioned stimulus-response pairs? In other words, was the increase which resulted from generalization as great as the increase which resulted from direct conditioning?

The stimuli nearest to each other were arbitrarily chosen for the comparisons, so that the smallest stimulus was compared with the second smallest, the lightest with the second lightest, etc.

In Exp. A the increase in conditioned responses to the smallest stimlus was not significantly different than the increase in appropriate responses to the second smallest stimulus. However, the increase in conditioned responses to the largest stimulus was significantly greater than the increase in appropriate responses to the second largest stimulus. In Exp. B, both tests of significance failed to yield a significant difference between conditioned and nonconditioned stimulus-response pairs. Table 5 summarizes the Wilcoxon significance tests.

TABLE 5

T Scores of Differences between Conditioned and Nonconditioned Responses in Postconditioning Increases to Appropriate Stimuli

Stimuli	*T* Found	*T* Needed[a]	
		.05	.01
Smallest—2nd Smallest	48.0	30	20
Largest—2nd Largest	0.0	40	28
Lightest—2nd Lightest	46.5	35	23
Darkest—2nd Darkest	32.0	21	13

[a] Probability of significance increases with decrease in *T*.

Discussion

This experiment indicates that discrete response choices can be used as though they are points on a continuum if they are presented in a way whereby they can be associated with each other via aspects of presentation. This is evidenced by the fact that the choices of 37 of 40 *Ss* in this study were predicted by a frame of reference which treated the discrete choices as belonging to a continuum.

The results, in supporting the hypothesis, indicate that stimulus-response generalization can take place with discrete response choices which are associated with each other via the presentation attributes. Extensions of generalization theory to situations involving discrete response choices would consequently appear tenable. A further implication is that *Ss* may learn more stimulus-response relationships than are directly conditioned, given ordered stimuli, and responses placed on a continuum. More specifically, compromise responses can be predicted to stimuli between conditioned stimuli.

An outcome of the use of discrete responses in this experiment has been the indication that it is not necessarily true that the conditioned response occurs most frequently to the variations of the stimulus to which the response has been conditioned. The availability of appropriate alternate responses can reverse the expected frequencies, so that the conditioned response decreases and is replaced by a nonconditioned response.

An attempt was made to compare the relative strengths of the conditioned and nonconditioned stimulus-

response pairs, in terms of amount of increased responding. A total of four comparisons were made comparing response increases due to generalization with response increases directly due to conditioning. Three of them yielded no significant differences, whereas one comparison, in Exp. A, indicated that the conditioned response increased to a significantly greater degree than the nonconditioned response. Although this leaves the question of relative strength unanswered, the indication is that the increase in responses resulting from generalization, under the conditions of this experiment, can, at times, be as marked as the directly conditioned increase.

Summary

Forty college students were tested for stimulus-response generalization, with an examination of discrete response choice variation, rather than an examination of the variations in a more limited single response class. The discrete response choices had been placed on a continuum via stimulus attributes involved in the presentation of response choices to *Ss*.

Only the ends of the stimulus and response continua had been conditioned, but it was predicted that intervening responses would be more frequently evoked by intervening stimuli following conditioning. The hypothesis was supported by the results. It was also found that the conditioned end responses decreased in frequency to the intervening stimuli nearest to the conditioned stimuli, when both the intervening and end responses were present as choices following conditioning of the end points.

References

Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

Siegel, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

(Received July 10, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 30–35

THE INFLUENCE OF THE READY SIGNAL AND UNPAIRED UCS PRESENTATIONS ON EYELID CONDITIONING

WALLACE R. McALLISTER AND DOROTHY E. McALLISTER

Syracuse University

Evidence has been presented (Kimble, Mann, & Dufort, 1955) indicating that, following a minimal number of paired CS-UCS trials, presentations of the UCS alone will increase eyelid conditioning at least as well as continued CS-UCS pairings. An attempt to duplicate this finding by Goodrich, Ross, and Wagner (1957) was unsuccessful. Since, however, the procedures of the two investigations deviated in many respects from each other, the discrepancy in results may be attributed to any one or more of the differing factors. One of these factors was the use of a ready signal: Kimble et al. did not use one; Goodrich et al. did. It has recently been shown (McAllister & McAllister, in press) that eyelid conditioning is affected markedly by this variable, the absence of a ready signal leading to better conditioning. Since the basis for the effect is not well understood, the ready signal seems a likely candidate as the variable causing the difference in results. The purposes of the present investigation were (*a*) to attempt to obtain the results of the Kimble et al. experiment by duplicating their procedures as closely as possible with the available apparatus, and (*b*) to test the hypothesis that the failure of Goodrich et al. to obtain the results of Kimble et al. was due to their use of a ready signal. Recently Dufort and Kimble (1958) have independently tested the same hypothesis but with results contrary to those found in this experiment.

METHOD

Subjects and design.—The *S*s were 90 students, 45 men and 45 women, from an introductory course in psychology who participated in the experiment in partial fulfillment of a course requirement. Within each sex, *S*s were assigned in order of appearance to one of three groups in accordance with a random order established beforehand. Each group thus consisted of 15 men and 15 women.

Two groups of *S*s were given 60 trials without the use of a ready signal. Group A received paired presentations of the CS and the UCS on Trials 1–20 and 41–60 and the UCS alone on Trials 21–40. Group C received paired presentations of the CS and UCS on all 60 trials. A comparison of the changes in performance of these groups from Trials 1–20 to Trials 41–60 would indicate how conditioning was affected by the different treatments given during the intervening trials. On the basis of the Kimble et al. study, equal increases in performances would be expected. Another group, Group B, was run under the same conditions as Group A except that a ready signal was given on each trial. If the use of a ready signal was the reason Goodrich et al. were unable to obtain the results of Kimble et al., it would be expected that an increase in conditioning from Trials 1–20 to Trials 41–60 would be shown by Group A but not by Group B.

The data obtained from an additional 12 *S*s were discarded for various reasons: 3 for *E*'s error, 2 for apparatus failure, 2 for unreadable records due to excessive random blinking, and 5 for meeting the voluntary conditioning criterion established beforehand. The *S*s were considered to be voluntary conditioners if over half of their CRs on Trials 1–20 and 41–60 occurred between 200 and 300 msec. after the presentation of the CS. This criterion was established in accordance with generally accepted procedures (Spence & Taylor, 1951). Of the 5 voluntary *S*s, 4 were women, 3 from Group C and 1 from Group B. The male voluntary *S* was from Group A.

Apparatus and procedure.—The CS was a 1000-cycle tone at a sensation level of approximately 55 db, generated by a Hewlett-Packard audio oscillator, Model 200 AB, presented through a General Electric speaker, Model S-1201. The UCS was a puff of air of approximately 4.5 lb./sq. in., delivered to the cornea of the right eye through a nozzle 1 mm. in diameter. The puff was generated by an air system consisting of an air compressor, a storage tank, a pressure-reducing valve, and a solenoid-operated air valve. The duration of the CS was 250 msec.; of the UCS, 100 msec., presented 250 msec. following the termination of the CS. Since there was a 40-msec. lag in the delivery of the air, the CS-UCS interval was 540 msec. Hunter interval timers controlled the onsets and durations of the stimuli.

A graphical system was used to record the eyeblink response. It consisted of a microtorque potentiometer connected mechanically to the eyelid of *S*, a Hunter eyeblink amplifier, an Edin 8002 penmotor, and a Hunter paper puller.

Each *S* was seated in a moderately illuminated (3.2 ft-c) soundtreated room in an armchair equipped with a head rest. The *E* administered the conditioning trials from an adjacent soundtreated room in which the apparatus controls were located.

The instructions read to *S* before the experiment were essentially the same as those used by Kimble et al. (1955, p. 408). Minor changes were necessitated by differences in apparatus. The instructions were intended to induce a neutral set in *S* and withheld the nature of the experiment by informing *S* that it was concerned with the reaction of the human eye to different kinds of stimulation, i.e., tone and puff.

The group which received the ready signal was instructed that they would be told "Ready" over the intercommunication system between the two rooms shortly before the stimuli were presented. Actually, the ready signal was given either 3, 4, or 5 sec. prior to the administration of the CS, the interval varying in an unsystematic manner from trial to trial. The *S* was also required throughout the experiment to look toward a gray cardboard, 6 in. high by 16 in. long, affixed to a shield at about eye level and approximately 24 in. in front of him. This procedure tended to stabilize the baseline of the recording.

Prior to the conditioning trials, 10 preliminary presentations of the CS alone were given at 15-sec. intervals. Any *S* who gave six or more responses classified as CRs in the 10 preliminary trials was to have been discarded. However, no *S* met this criterion. Following the preliminary trials, 60 trials were administered with a fixed 15-sec. intertrial interval under one of the three experimental conditions.

Results

In Fig. 1 are presented the mean percentages of CRs as a function of blocks of trials for each of the groups. Separate curves for men and women are not shown since the same relationships were found with each sex, although in every instance the women's curve was displaced upward relative to the men's.

Inspection of the curves for Groups A and C indicates that the results of Kimble et al. were not duplicated in this investigation. The gains in percentage of CRs for these groups from Trials 11–20 to Trials 41–50 were analyzed with a simple factorial analysis of variance. One factor was

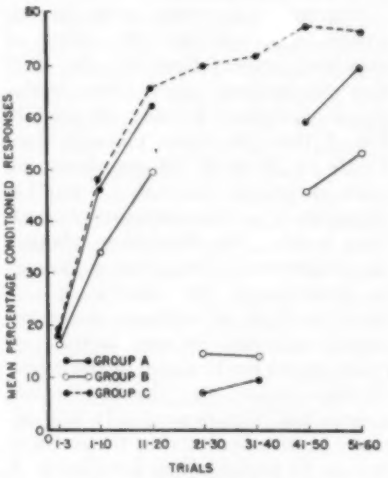

FIG. 1. Mean percentage of CRs as a function of blocks of trials for the several groups, sexes combined. Groups A and C received no ready signal while Group B did. Groups A and B had the UCS presented alone over Trials 21–40 while Group C had the CS paired with the UCS on these trials.

the paired or unpaired presentation of the UCS in Trials 21–40; the other was sex.[1] The mean gains for the factors were: men, 3.0; women, 5.9; paired, 11.7; unpaired, −2.7. Neither the sex factor nor the interaction effect was significant ($F < 1$ in both cases). The paired-unpaired UCS factor was, however, significant beyond the 1% level ($F = 7.48$, 1 and 56 *df*). To determine whether the absolute change in performance over Trials 21–40 was significantly different from zero for each of the groups, t tests were computed. For the paired group, the gain in performance was significant beyond the 1% level ($t = 3.372$, 29 *df*) while for the unpaired group the change in performance (a decrement) was not significant ($t = .696$, 29 *df*). These results indicate that gains in conditioning occurred only when the CS and UCS were paired.

The second question under investigation was whether the effect of unpaired presentations of the UCS was dependent upon the ready-signal variable. It may be seen in Fig. 1 that the mean changes from Trials 11–20 to 41–50 are about the same in Groups A and B, 2.7 and 3.4, respectively, and are decrements rather than gains. To determine whether the difference in these mean changes in performance was reliable, a factorial analysis of variance was computed with sex as one factor and ready vs. no ready signal as the other. Neither factor, nor the interaction, approached significance. It has already been indicated that the absolute change in performance for Group A was not significant. The same result was obtained with a comparable analysis made on the performance of Group B ($t = .969$, 29 *df*). Therefore, it can be concluded that presentation of the UCS alone brought about no significant improvement in performance regardless of whether or not a ready signal was used.

The above evidence indicates that conditioning increased only with paired CS-UCS presentations. Therefore, if the unpaired puffs had no effect upon performance, it would be expected that the level of conditioning of Groups A and C would be equal after the same number of paired trials. As may be seen in the figure, however, on the third block of paired trials the mean performance of Group A (59.5% on Trials 41–50) is lower than that of Group C (69.7% on Trials 21–30). This difference, which is in agreement with the findings of Goodrich et al., suggests a decremental effect of the unpaired puffs. However, in neither experiment was the difference between the groups statistically significant.[2] Despite this lack of significance, the consistency in the direction of the difference and the observations made by *E*s during the present experiment suggested an examination of the individual trials. An inspection of the records indicated that the lowered performance seemed to be largely the result of one trial, Trial 41, which was the first trial the CS was paired with the UCS following the presentation of the puffs

[1] Since sex was controlled in the assignment of *S*s to conditions in order to increase the precision of the experiment, it was used as a control factor in all analyses of variance. For each analysis, the assumption of homogeneity of variance was tested and found to be tenable.

[2] In the present experiment a factorial analysis of variance with Groups and Sex as the factors indicated that the difference in performance between the Groups was significant at only about the 20% level ($F = 1.68$, 1/56 *df*). The interaction was not significant ($F < 1$), but the women gave a significantly higher percentage of CRs ($F = 4.72$, 1/56 *df*) than did the men (73.1% and 56.1%, respectively).

TABLE 1

MEAN PERCENTAGE OF CONDITIONED RESPONSES ON INDICATED TRIALS FOR THE THREE GROUPS

Group	Trials Compared				
	(1)	(2)	(3)	(4)	(5)
	11–20	20	41 (A & B), 21 (C)	42 (A & B) 22 (C)	41–50 (A & B) 21–30 (C)
A	62	63	17	67	60
B	49	30	10	52	46
C	66	60	67	63	70

alone. The pertinent data are summarized in Table 1. A comparison of Columns 2 and 3 indicates a large decrement in performance from Trial 20 to the next paired trial for Groups A and B but not for C. In fact, the percentage of CRs of Groups A and B remains at about the same level as on their puff-alone trials as is shown in the figure. Consideration of Column 4, however, indicates that recovery occurred by the second paired trial following the UCS-alone presentations. As can be judged from Column 5, that level was maintained for the remainder of the block of trials. A test of correlated proportions (McNemar, 1949, p. 80 f.) indicated that the shift in proportion of CRs from Trial 20 to the next paired trial for Group A was significantly different from that of Group C (CR = 3.25). The same test applied to the shift in performance from Trial 20 to the second paired trial following was not significant (CR = .08). The same two analyses made between the performances of Groups A and B yielded no significant differences (CR = 1.53 and 1.06, respectively). It should be noted that Group B's performance on Trial 20 is not representative. As shown in Column 1, the more stable measure at that stage of practice was much higher. However, if the performance on Trial 20 had been normal, the interpretation of the statistical test would not have been changed. These findings indicate that the effect of the unpaired puffs was to produce at least a temporary decrement regardless of whether or not a ready signal was used.

Evidence concerning the effect of the ready signal upon conditioning is also provided by the present data. The performance of Group B was compared with that of Groups A and C combined over Trials 1–20 and with that of Group A over Trials 41–60. Both statistical analyses allowed the conclusions to be drawn that the use of a ready signal lead to poorer conditioning and that women condition better than men. The effect found with the ready signal confirms results previously reported (Dufort & Kimble, 1958; McAllister & McAllister, in press). The latter investigators, in addition, report similar tendencies with the sex factor.

DISCUSSION

When a ready signal has been used, previous investigators have been consistent in reporting no improvement in eyelid conditioning as a result of the interpolation of a series of unpaired puffs (Dufort & Kimble, 1958; Goodrich

et al., 1957). The present results are also in agreement. With no ready signal, the interpolated UCSs have been reported (Dufort & Kimble, 1958; Kimble et al., 1955) to be as effective in increasing eyelid conditioning as paired CS-UCS trials. The present results are in conflict with these latter findings even though the experimental conditions closely duplicated those of Kimble et al.[3] The inconsistency must be a result of some procedural difference between the investigations. Obvious differences such as the use of an auditory rather than a visual CS and the dissimilarities of the physical environments do not seem to be likely sources of the discrepancy. Other differing factors (criteria for counting the CR, recording equipment, etc.) similarly do not seem to be the crucial variables since Dufort and Kimble found that the UCS alone improved conditioning when such factors were similar to the present ones. Conceivably, some complex interaction of variables is involved. At any rate, it does not seem possible at present to point to any known procedural difference as being responsible for the diverse results.

It might be expected that the use of a fixed intertrial interval in this investigation, employed to duplicate the procedures of Kimble et al., would lead to temporal conditioning. In fact, Goodrich et al. suggested this as a possible explanation of the improvement in conditioning following unpaired puffs originally reported by Kimble et al. Since Dufort and Kimble later found the effect when using a variable intertrial interval, the explanation does not seem to be a plausible one. Nevertheless, the possibility that temporal conditioning may have contributed to the increased performance found in the Kimble et al. study is not ruled out. The present study, which used a graphical method of recording, provides information pertinent to this question. If temporal conditioning was occurring, it would be expected that some increase in the number of CRs would occur during the 20 unpaired trials. The lack of improvement over the puff-alone trials, as can be seen in Fig. 1, does not support such a notion. The level of performance obtained on these trials can probably be accounted for most parsimoniously as random blinks occurring in the interval in which CRs were counted. This explanation is consonant with the results of an examination of the records which indicated that the form of most of the responses occurring on the puff-alone trials was unlike the typical CR on paired trials.

The finding that there was a significant decrement in conditioning on the first trial following the interpolation of unpaired puffs (relative to a group which did not receive the unpaired puffs) is difficult to interpret. Since the level of performance prior to the puff-alone trials was at a high level, the excitatory tendency must have been well above threshold. The interpolated puffs would be expected to maintain the drive level (Spence, Haggard, & Ross, 1958); therefore, when the CS was given, the CR should have occurred. It might be suggested that following the series of unpaired puffs the *Ss* might not attend to the first presentation of the CS. Such an explanation seems less likely when it is noted that the group given the ready signal, and thus having their attention drawn to the impending presentation of the trial, showed the same decrement as the group without the ready signal. Goodrich et al. suggested that the lower level of conditioning they obtained on the block of trials following the unpaired puffs, relative to a group with an equal number of paired trials, might be the result of adaptation to the UCS. To the extent that the observed decrement is attributable to the first trial following the puff-alone trials, adaptation does not seem to be an adequate explanation since its effect would not be expected to disappear with one pairing of the CS with the UCS.

[3] Recently, Goodrich, Ross, and Wagner (1959) have also reported failure to obtain an increase in conditioning due to unpaired puffs when no ready signal was used.

SUMMARY

Three groups of *Ss*, 15 men and 15 women in each, were given 60 trials in an eyelid conditioning experiment. For one group (the control group) the CS was paired with the UCS on every trial; for the other two groups, only the UCS was presented on Trials 21–40. One of the latter two groups was given a ready signal before each trial while the other group and the control group were not. The results indicated that the unpaired trials did not improve conditioning regardless of whether or not a ready signal was used. In fact, a marked decrement, which persisted for only one trial, occurred following the puff-alone trials for both groups. Confirming previous results, the groups without the ready signal conditioned better than those with it, and the women conditioned better than the men.

REFERENCES

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

GOODRICH, K. P., ROSS, L. E., & WAGNER, A. R. Performance in eyelid conditioning following interpolated presentations of the UCS. *J. exp. Psychol.*, 1957, **53**, 214–217.

GOODRICH, K. P., ROSS, L. E., & WAGNER, A. R. Supplementary report: Effect of interpolated UCS trials in eyelid conditioning without a ready signal. *J. exp. Psychol.*, 1959, **58**, 319–320.

KIMBLE, G. A., MANN, L. I., & DUFORT, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407–417.

MCALLISTER, W. R., & MCALLISTER, D. E. The ready signal in eyelid conditioning. *Amer. J. Psychol.*, in press.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received July 20, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 36–40

REVERSAL LEARNING UNDER SINGLE STIMULUS PRESENTATION

DAVID BIRCH, JAMES R. ISON, AND SALLY E. SPERLING

University of Michigan

In simple form the procedure employed in the reversal problem involves two stages. In the initial stage, approaching responses to one of two discriminanda are reinforced while approaching responses to the other are not reinforced. In a subsequent stage, the relation of reinforcement to stimuli is reversed so that approaching the previously reinforced cue is now nonreinforced and vice versa.

Reid (1953), Pubols (1956), and Capaldi and Stevenson (1957) report data which indicate that rats which have been provided extended overtraining on an initial Black-White simultaneous discrimination learning problem require fewer trials to criterion on the subsequent reversal problem than *S*s which are reversed at criterion performance on the original problem. An analysis of the reversal problem at a molar level points up two factors, either or both of which could operate to produce the obtained results. The difference between the criterion and overlearning *S*s during reversal training could arise from differential rates of extinction to the previously positive cue and/or from differential rates of acquisition to the newly reinforced cue for the two groups.

In order to investigate the importance of these two factors it is necessary that separate measures of response tendency strength be obtained for the two discriminanda. Simultaneous discrimination learning is not well adapted for this purpose since a given choice response is determined by the combination and interaction of the approaching tendencies to a complex of stimuli including both the two discriminanda and the directional cues of the situation.

In the present experiment single stimulus presentation in a straight runway apparatus with response speed as the measure of response tendency strength is employed to investigate reversal learning. Grice (1948), Logan (1952), and Birch (1955) have presented evidence on the relationship between the single stimulus response speed measure and the choice measure.

For the purposes of the present study the single stimulus presentation procedure has three clear advantages over a simultaneous choice situation as a method of investigation: (*a*) separate measures of response strength to the two discriminanda are obtained, (*b*) the possibility of differentially strong position habits is eliminated, and (*c*) the importance of complex receptor-orienting acts is minimized. The third advantage is of special interest since both Reid (1953) and Pubols (1956) suggest that their reversal results may be accounted for in terms of a response of discriminating, which is learned concurrently with the choice response. By assuming that the response of discriminating transfers to the reversal problem with a facilitating effect and that this response exhibits greater resistance to extinction for the overlearning *S*s, the obtained results are derived.

The response of discriminating is described by Reid as a response "to

a set of stimuli within the total stimulus complex—in this case the black and white stimulus cards" (1953, p. 110), and is identified by Pubols as "orientation of the eyes toward differential cues" (1956, p. 248) in the case of visual discriminations. The importance of the response of discriminating as an explanatory device appears to be related to the insuring of advantageous reception of the two discriminanda presented simultaneously on each trial. The utility of such orienting acts should be minimized, if not eliminated, by the single stimulus procedure which presents *S* with only one of the two cues on each trial.

PROCEDURE

Subjects.—Twenty-five male hooded rats, aged 75 to 90 days at the beginning of the experiment were obtained from the colony maintained by the Psychology Department of the University of Michigan. Two animals were eliminated because of failure to eat during pretraining and two because of failure to run on the first day in the apparatus.

Apparatus.—The apparatus consisted of two elevated platforms enclosed in a cubical compartment, 8 ft. to a side, constructed from ½-in. beaverboard on a pine frame. A doorway 2 ft. wide and 8 ft. high and covered with a black curtain provided *E* with access to the cube. The compartment was divided by a T-shaped partition, also 8 ft. high, into a starting platform section, a stimulus platform section, and an entranceway. Except for the floor of the stimulus platform section, which was of midgray floor tile, the entire apparatus was painted flat black.

The starting platform, 26 in. from the floor, was 12 in. long and 3 in. wide. Its semicircular terminus protruded through a portal, 3 in. wide and 4 in. high, located in the center of the T-shaped partition. The stimulus platform, 12 in. long and 4 in. wide, was placed in line with the starting platform but ½ in. lower and separated from it by a 1-in. gap. A small unpainted aluminum foodcup could be attached to the terminal end of the stimulus platform. Black and white stimulus platforms were used and could be changed on the apparatus by an arrangement of snap fasteners.

A skylight of translucent plastic, 2 ft. square, was located in the cube directly over the stimulus platform. Attached to the ceiling of the outer room directly over the skylight was a 150-w. bulb in circuit with a rheostat set to give a constant illumination of 5.29 ft.-L., as measured by a standard MacBeth illuminometer, to the stimulus platform section when the circuit was closed. A 7-w. bulb attached to the T partition 3 ft. above the starting platform illuminated the starting platform section. An infrared light beam, obtained from two photoelectric cells placed outside the cube, was directed 1 in. over the starting platform, ½ in. from the partition. When this beam was broken by *S*, (*a*) the light in the starting platform chamber went out, (*b*) the light above the stimulus platform chamber skylight went on to illuminate the stimulus platform, and (*c*) a General Electric standard electronic timer, calibrated in .01 sec., was started. When *S* stepped off the starting platform, a normally off microswitch placed under that platform was released to stop the clock. A black door normally raised 1 ft. above the starting platform was lowered by E_1 from the outer room to prevent retracing. The E_2 recorded the time elapsed between the stimulus platform presentation and *S*'s approach.

Preliminary training.—The pretraining period extended over 14 days. The *S*s were maintained on a 22-hr. feeding schedule in which they ate 8 gm. of Pard dog meal during a 1-hr. feeding period. Beginning on Day 4 and continuing through Day 9, *S*s were handled individually for 5 min. daily. On Days 10 through 13 they were placed on an elevated runway of natural pine and allowed to eat from attached foodcups. On Day 14 they were given one rewarded trial in the training apparatus with the positive stimulus platform in place.

Discrimination, overlearning, and reversal training.—The *S*s were run in squads of four or five with a minimum intertrial interval of 5 min., during which they were housed in individual carrying cages. In discrimination training the white platform had attached to it a foodcup baited with two .05-gm. pellets and *S* was removed immediately after eating. The *S* was detained on the black platform for 15 sec. with no foodcup present.

The *S* was given 11 trials each day with the first always to the positive white platform. The first trial was considered as warm-up and is not included in any comparisons between groups or in considering performance to the criterion. The following 10 trials were

TABLE 1

NUMBER OF DAYS TO CRITERION ON THE ORIGINAL AND REVERSAL PROBLEMS FOR THE CRITERION AND OVERLEARNING GROUPS

Group	*N*	Original		Reversal	
		Mean	*SD*	Mean	*SD*
Overlearning	11	4.7	2.0	3.6	1.1
Criterion	10	4.3	1.6	5.3	1.6

always 5 positive and 5 negative with the sequence determined by randomly selected Gellerman (1933) orders. The criterion of learning for the discrimination problem was that all five responses to the white platform be faster than the fastest response to the black within a single day.

Each *S* was assigned at random to either the Criterion Group ($N = 10$) or the Overlearning Group ($N = 11$) before discrimination training was begun. On the day following criterion performance on the assigned task, the *S*s of the Criterion Group were transferred to the reversal problem with black positive and white negative. The *S*s in the Overlearning Group continued on the original task for 12 additional days following criterion performance after which they were transferred to the reversal problem. The first trial of the first day of reversal training was a reinforced trial to the old positive stimulus so that reversal training proper started with the second trial of that day.

RESULTS

Table 1 presents the data for the Overlearning and Criterion Groups on the original and reversal problems in terms of the gross measure of the mean number of days to criterion. The criterion day for an *S* was defined to be the first day in which all five latencies to the positive stimulus were less than the smallest latency to the negative.

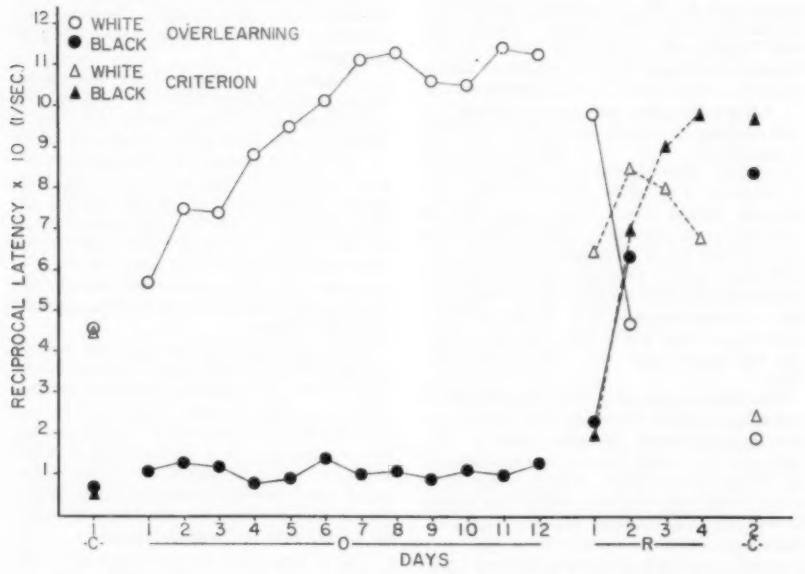

FIG. 1. Mean response speed to the white and black goal platforms for the Criterion and Overlearning Groups under original and reversal training. Points are given for the original problem criterion day (C1), for overlearning (O1–O12), for the initial segment of reversal training (R1–R4), and for the reversal problem criterion day (C2).

TABLE 2

ANALYSES OF VARIANCE OF THE RESPONSE SPEEDS FOR EXTINCTION AND ACQUISITION UNDER REVERSAL TRAINING

Source	*df*	Extinction		Acquisition	
		MS	*F*	*MS*	*F*
Between *Ss*	20	2909.36		748.56	
Groups (G)	1	6311.39	2.312	1.55	<1
Error (b)	19	2730.30		787.88	
Within *Ss*	42	1177.76		1192.45	
Days (D)	2	1643.53	3.548*	19,464.64	69.417**
G × D	2	14,288.71	30.848**	249.28	<1
Error (w)	38	463.20		280.40	
Total	62				

* $P = .05$.
** $P = .001$.

The two groups show no statistically significant difference in the mean number of days to criterion on the original problem ($t = 0.48$, $df = 19$). On the reversal problem, however, the Overlearning Group is superior to the Criterion Group ($t = 2.74$, $df = 19$, $P < .02$).

The source of the difference in performance on the reversal problem can be found by an examination of the mean response speeds of the two groups to the positive and negative stimuli. Figure 1 presents the response speed data (10 times the reciprocal latency) for those days in which all *Ss* of a group can be entered into the calculation of the mean.

The Overlearning Group, responding considerably faster than the Criterion Group to the positive stimulus at the conclusion of training on the original problem, shows a monotonically decreasing extinction function to that stimulus under the reversal condition. The Criterion Group, in contrast, extinguishes more slowly and even appears to increase the speed of response to the formerly positive stimulus during the initial stage of the reversal problem. The course of acquisition to the newly positive stimulus is similar for the two groups.

Separate analyses of variance were carried out on the extinction and acquisition performance of the two groups using the response speed measure. The two analyses are summarized in Table 2.

Both analyses include the response speeds for the last day of training on the original problem and the first two days of reversal training for the two groups. Of particular interest is the Groups × Days interaction in each case. This interaction is significant beyond the .001 level for the extinction to the formerly positive stimulus but negligible for the acquisition to the newly positive stimulus.

DISCUSSION

The results show clearly that the basic finding of the reversal problem can be obtained under single-stimulus presentation. That is, animals provided overlearning on the initial problem require fewer trials to criterion on the subsequent problem than do nonoverlearning animals. Furthermore, the difference between the two groups appears to be located in their differential rates of extinction to the previously reinforced

cue. No difference in rate of acquisition to the newly positive cue was found in the present data.

The conditions imposed by the single-stimulus procedure effectively rule out interpretations of the results in terms of differential response tendency strengths to position cues or in terms of responses of discriminating simultaneously presented discriminanda. This is not to be construed as implying that either or both of these factors are not operative in those theoretically complex situations where the organism makes a choice in the presence of a variety of simultaneously presented cues. Rather, the implication is that further experimentation on the empirical relation between resistance to extinction and the number of acquisition trials is needed.

The present data indicate that, at least for the conditions of this experiment, resistance to extinction is not a monotonically increasing function of the number of acquisition trials. In fact, there is the suggestion that the group with fewer acquisition trials not only extinguished more slowly but also that their response speed increased during the initial period of the extinction conditions.

The possibility that differential rates of extinction for the Criterion and Overlearning Groups may account for the reversal problem results is recognized by Capaldi and Stevenson (1957) and is described by them in terms of *S*'s capacities to discriminate schedules of reinforcement. The present study, while supporting their empirical hypothesis about extinction, does not bear directly on the tenability of their proposed theoretical justification for this hypothesis.

Summary

The experiment investigated the discrimination reversal problem in rats under single-stimulus presentation with response speed as the measure of performance. Twenty-one male hooded rats under 22-hr. food deprivation were reinforced with food for approaching a white platform and nonreinforced for approaching a black. Five trials to each stimulus, arranged in randomly selected Gellerman (1933) orders, were given each day preceded by a single "warm-up" trial to the positive stimulus. Training on the original problem was continued until all five latencies to the negative were larger than the greatest latency to the positive. The Criterion Group ($N = 10$) was then placed into reversal training while the Overlearning Group ($N = 11$) was continued for 12 additional days training before reversal.

The two groups show no difference in days to criterion on the original problem but differ significantly on this measure for the reversal problem with the Overlearning Group superior to the Criterion Group. This result replicates with single stimulus presentation the reversal problem findings for simultaneous stimulus presentation.

Examination of the response speeds to the positive and negative stimuli indicates that the reversal problem difference for the groups may be attributed to the differential rates of extinction to the previously positive stimulus rather than to acquisition to the newly positive cue.

References

Birch, D. Discrimination learning as a function of the ratio of nonreinforced to reinforced trials. *J. comp. physiol. Psychol.*, 1955, **48**, 371–374.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 207–208.

Grice, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.

Logan, F. A. Three estimates of differential excitatory tendency. *Psychol. Rev.*, 1952, **59**, 300–307.

Pubols, B. H. The facilitation of visual and spatial discrimination by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

(Received July 22, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 41–49

SUCCESSIVE ACQUISITIONS AND EXTINCTIONS AS RELATED TO PERCENTAGE OF REINFORCEMENT [1]

GLEN D. JENSEN AND JOHN W. COTTON

Northwestern University

It is considered a well established fact (Jenkins & Stanley, 1950) that animals which are rewarded 50% of the time during training show greater resistance to extinction than animals that receive 100% reinforcement during training. Granting the empirical truth of this generalization, one may yet ask for a great deal of more specific information about the effect of 50% reinforcement. One might ask, e.g., whether resistance to extinction will be strongest if alternate trials are rewarded, if trials are rewarded in a random sequence, or if blocks of trials are rewarded (and there are many possible block-sizes if the training sequence is extensive).

The present paper reports a parametric study of the size of blocks of rewarded trials within a constant 50% reinforcement ratio. A very extensive (270 trials) training regimen was employed to allow variation of block size up to 30 trials while allowing at least nine transitions from rewarded to nonrewarded blocks of trials. Because 30 trials were given on each day of training, it can be seen that the group which received reward and nonreward in blocks of 30 trials fits the paradigm of a successive acquisition and extinction study. In addition to the partially reinforced groups a continuously reinforced group was run to serve as a baseline in determining partial reinforcement effects.

[1] This experiment was performed under a National Science Foundation grant (NSF G-2672).

METHOD

Subjects.—The *Ss* whose data are reported were 140 naive male albino rats from 59 to 103 days in age, with a mean of 78.4 days. The *Ss* ranged in weight on Experimental Day 1 from 191 to 265 gm., with a mean of 215 gm. Nine *Ss* were discarded because of procedural errors, two because they failed to learn the running response, and one because it would not eat the reward pellets. The *Ss* were assigned at random to seven groups of 20 *Ss* each, subject to the following restriction: The first seven *Ss* included one member of each group, the next seven *Ss* included one member of each group, and so on. Animals discarded in a set of seven *Ss* were replaced at the end of the experiment.

Apparatus.—A runway, previously employed by Lewis (1956) was used. Inside width was 4 in., inside height was 9 in., and length from the end of the startbox to the beginning of the goalbox was 4 ft. The duration of each response was defined as the time that elapsed from the moment the startbox door was raised until *S* placed weight on the goalbox floor, depressing a microswitch. A Standard Electric timer measured the duration of a response to the nearest .01 sec.

Procedure.—Prior to Experimental Day 1 all *Ss* received a 7-day schedule of deprivation and taming. During each of these seven days *S* received 10 gm. of finely ground Purina lab checkers 1 hr. later than the time of day when *S* was later to be run. Usually *S* finished this ration within 1 hr. On those few occasions (usually on the first two days of scheduling) when *S* did not finish the ration within 1 hr., it was allowed a second hour in which to finish it. No *S* took more than 2 hr. to finish the ration. This feeding procedure and ration size was chosen because it was desired to keep hunger drive as constant as possible, and preliminary experimentation had shown that 10 gm. of such food is a necessary and sufficient daily ration size to keep body weight approximately stable over two weeks of scheduling. During

the 10 days of experimentation *Ss* gained, on the average, 5.85 gm. or about 2 to 3% of their starting weight.

During each of the seven days *S* was handled by *E* for 5 min. and was allowed to eat six 45-mgm. pellets from a small cup later to hold the reward pellet in the goalbox.

During Experimental Day 1 each *S* received 30 acquisition trials in the runway. On each trial *S* received a 45 mgm. pellet. The *S* was left in the goalbox for about 10 sec. on each trial and about 15 sec. elapsed between trials during which time *S* waited in its home cage. If *S* had not reached the goalbox in 2 min., it was manually placed in the goalbox and allowed to eat the reward pellet. On such trials 120 sec. was arbitrarily recorded as the response time.

Differential treatment of the seven groups consisted in the pattern of rewarded and nonrewarded trials administered during Experimental Days 2 through 9. During each of these days each *S* was given 30 trials in the runway. The *Ss* in one group (henceforth called Group A) were rewarded on every one of the 240 trials of Days 2 through 9. Rats in Group 30E-30A received 30 rewarded trials on Days 3, 5, 7, and 9 but received 30 nonrewarded trials on Days 2, 4, 6, and 8. The *Ss* in Group 15E-15A received 15 nonrewarded trials at the beginning of each daily session followed by 15 rewarded trials. The *Ss* in the 10E-10A group received 10 nonrewarded trials at the beginning of Day 2 followed by 10 rewarded trials followed in turn by 10 nonrewarded trials. On subsequent days for this group an alternating extinction–acquisition sequence was continued so that Days 2, 4, 6, and 8 started with 10 nonrewarded trials, whereas Days 3, 5, 7, and 9 started with 10 rewarded trials. On each day the 5E-5A *Ss* received 5 nonrewarded (E) trials, then 5 rewarded (A) trials, then 5 nonrewarded trials, and so on so that the daily sequence was 5E-5A-5E-5A-5E-5A. One group of *Ss* (R) received 15 rewarded and 15 nonrewarded trials on each of Days 2 through 9 according to a randomized schedule.[2] A final group (1E-1A) received 50% reinforcement on each of Days 2 through 9 according to an alternating schedule so arranged that the first trial of each day was nonrewarded and the last trial was rewarded.

Twenty-four hours after Day 9 each *S* received 30 nonrewarded trials. On every rewarded or nonrewarded trial during each of the 10 days of experimentation *S* was left in the goalbox for approximately 10 sec. after which he was placed in his home cage for about 15 sec. to await the next trial. The *S* was manually placed in the goalbox if he had not stopped the clock (entered the goalbox) within 2 min., and 2 min. was assigned as his running time. Twenty-four hours elapsed between each two successive experimental days except Days 5 and 6. Day 5 always came on a Friday and the schedules of the undergraduate *Es* necessitated leaving a 72-hr. interval between Days 5 and 6. During the weekend between Days 5 and 6, *Ss* were fed at the regular time.

All running-time scores were transformed to their reciprocals. These reciprocal values (henceforth termed "speeds") enter into all graphs and analyses.

Results

Partial reinforcement effects.—The course of running behavior over trials as a function of the experimental conditions is shown in Fig. 1. Because the data are so intertwined, the figure has been divided into two parts to facilitate rapid appraisal. Note that the 100% group (Group A) showed both a slower mean speed and a slower terminal speed for Day 10 (extinction) than did any other group, although the 30E-30A group was almost as slow. An analysis of variance using the mean speeds for all 30 trials of Day 10 yielded an F of 16.96 ($P < .001$, 6/133 *df*). The Tukey[3] procedure for comparing individual means in the analysis of variance revealed that Group A ran signifi-

[2] A total of 40 schedules were constructed by means of a table of random numbers with no restrictions except that 15 trials be rewarded. An individual schedule was constructed by reading the last two digits in a column of numbers until exactly 15 numbers between or including 01 and 30 were obtained: These 15 numbers designated the trials that were rewarded. With 40 random schedules it was possible to assign a different schedule to a specific *S* for each of Days 2 through 9. Further, exactly 4 rats received the same schedule on a given day.

[3] Tukey, J. W. The problem of multiple comparisons. Privately circulated monograph, 1953.

cantly slower than Groups 15E-15A, 1E-1A, R, and 5E-5A; Group 30E-30A ran slower than Groups 15E-15A, 1E-1A, R, and 5E-5A; Group 10E-10A ran slower than Groups 1E-1A, R, and 5E-5A; and Group 15E-15A ran slower than Groups R and 5E-5A. No other comparisons are significant at the .05 level. Thus, when the mean of all 30 extinction trials serves as the criterion, four groups (1E-1A, R, 5E-5A, and 15E-15A) show traditional partial reinforcement effects while two groups (10E-10A and 30E-30A) show no significant effect. The lack of a significant difference between Group 1E-1A and Group R fails to confirm the previous finding of Tyler, Wortz, and Bitterman (1953), although the direction of the difference is in line with their finding.

To test whether terminal level of extinction was a function of the experimental conditions, an analysis of variance was performed using the mean for each *S* of the last 5 trials of Day 10. The *F* obtained was 7.20, which is highly significant ($P < .001$, 6/133 *df*). The Tukey procedure revealed that Group A ran significantly slower than Groups 1E-1A, 15E-15A, R, and 5E-5A ($P < .05$) and almost significantly slower than Group 10E-10A; Group 30E-30A ran significantly slower than Groups 15E-15A, R, and 5E-5A ($P < .05$). No other comparisons are significant. Thus all groups except Group 30E-

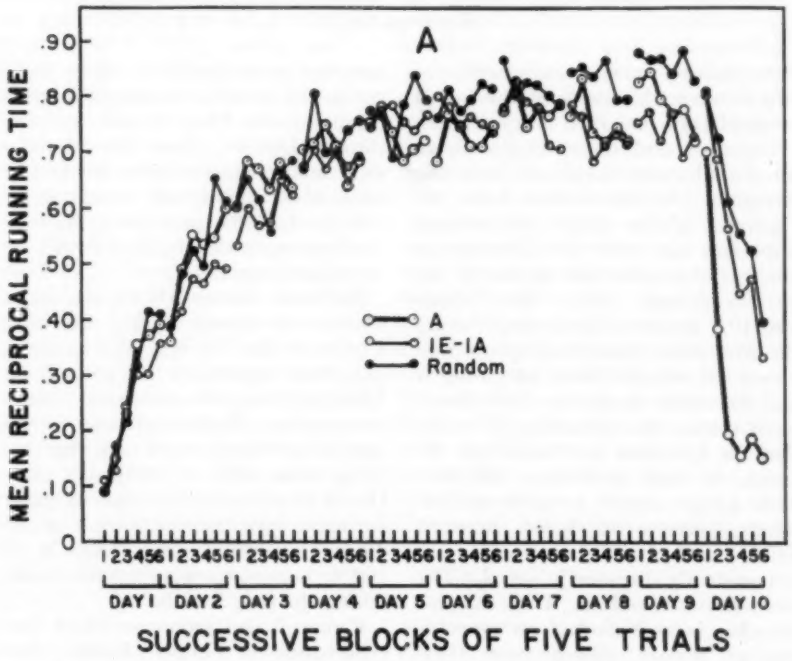
A
MEAN RECIPROCAL RUNNING TIME
.90
.80
.70
.60
.50
.40
.30
.20
.10
0
A
1E-1A
Random
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
DAY 1 DAY 2 DAY 3 DAY 4 DAY 5 DAY 6 DAY 7 DAY 8 DAY 9 DAY 10
SUCCESSIVE BLOCKS OF FIVE TRIALS

FIG. 1. Mean reciprocal running time in seconds as a function of trials. Trials are grouped in blocks of five. Day 10 is an extinction session for all groups.

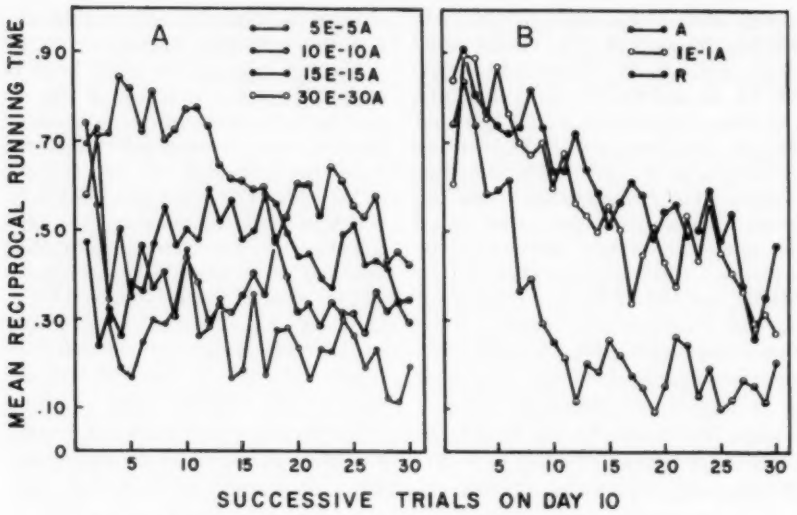

FIG. 2. Mean reciprocal running time (seconds) as a function of successive trials during Day 10.

30A showed partial reinforcement effects when the mean of the last 5 trials of Day 10 served as a criterion.

Figure 1 and the analyses above are not adequate to convey the whole flavor of the data from Day 10. Figure 2, giving single trial means, impresses one with the difference in *shape* of the extinction curves for the various groups. Note that Group 10E-10A shows a rapid drop during the first three trials to a point well above the terminal level for Group A and thereafter shows very little decrement during the remaining 27 trials. Perhaps the most curious of all the curves is that of Group 15E-15A. This group shows a more or less steady *increase* in speed between Trials 2 and 15 and thereafter a more or less steady decrease in speed. The increase in speed as Trial 15 approaches is as if *Ss* had an expectation of reward arising from their experience during the preceding 8 days of training. Figure 1 reveals a progressive tendency for these *Ss* to run faster as extinction trials progress beginning at Day 5 and carrying through Day 10. Even Days 3 and 4 show less decrement during the first 15 trials than one would expect if *Ss* had not had the previous experience of 15 nonrewarded trials followed by 15 rewarded trials.

Similarly Group 5E-5A shows an increase in speed during the first 4 trials of Day 10, very likely reflecting their experience of reward on Trial 6 during the previous 8 days of training. Single-trial data (mean speeds of 20 *Ss* for each trial) for this group show that on every day after Day 2 there was an increase in speed during at least the first 3 trials of the day and on three days (Days 4, 7, and 9) there was a progressive increase during the first 5 trials.

Figure 2 also shows an effect like expectation in Group 1E-1A. The first data point is far below the corresponding data point for Group A

(reflecting 8 days of experience of a nonrewarded first trial), and the second data point is well above the corresponding point for Group A and very far above the first-trial mean for Group 1E-1A. On Day 9 the mean speeds for all 15 rewarded trials of Group 1E-1A were well above the mean speeds for the 15 nonrewarded trials, forming a saw-toothed effect in confirmation of Tyler, Wortz, and Bitterman (1953) and indicating discrimination of rewarded from nonrewarded trials.

Acquisition and terminal level of performance.—Jenkins and Stanley (1950) conclude that partial reinforcement leads to slightly slower acquisition and maintenance at a slightly lower level than continuous reinforcement. This generalization was generated largely by studies in which only a 100% group and a 50% random reinforcement group were used and in which a much shorter training period was used than in the present study. From Fig. 1 it can be seen that if the present study had employed only two such groups (A and R) and if training had stopped after 90 trials, that generalization would have been more or less supported (although the mean speed for Day 2 is a trifle higher for Group R relative to Group A). However, Fig. 1 shows that as training progresses beyond the third day, Group A (100%) becomes progressively more and more inferior in mean speed to Group R (50%). Moreover, all groups except Group 1E-1A show a higher terminal level of performance than Group A if one uses the last trials of Day 9 as a criterion. Even Group 1E-1A shows a higher terminal level of performance than Group A if one considers the last few *rewarded* trials on Day 9 (Fig. 1 does not show this fact, but a graph of single-trial data would bring it out clearly). That the inferiority of Group A is not the result of selecting inferior *S*s for Group A is suggested by the fact that the mean speed of Group A for Day 1 was second only to that of Group R, with Group R being only .003 sec. faster in mean speed. An analysis of variance performed on the mean speeds for Day 1 yielded an F of only .22, indicating no appreciable differences in initial ability between groups.

In performing a statistical test of differences in terminal level of performance, a mean speed during the last 5 rewarded trials of Day 9 was first obtained for each *S* in all groups except Group R. For Group R the mean speed for the last 5 trials of Day 9 was used because *S*s in that group had no basis for discriminating between rewarded and nonrewarded trials before they were completed, except for the fact that a total of 15 of the trials had to be rewarded. A simple analysis of variance was performed on these scores. The F obtained was 3.88 which is highly significant ($P < .01$, 6/133 df). A subsequent Tukey analysis revealed that Group A was slower than all except Groups 1E-1A and 30E-30A at the .05 level or better, and Groups 1E-1A and 30E-30A were slower than Group A very nearly at the .05 level. No other comparisons were significant. The fact that the last data point on Day 9 for Group A is the lowest point of the day should mitigate conclusions drawn from this analysis. This point will be discussed further below.

Reinforcing effects of extinction.—The paradox implied by the subtitle will disappear if the following definitions are employed. Let a "reinforcement" be any object or event which follows closely the occurrence

of a response and more or less permanently increases response strength. Let "extinction" refer to *E*'s operation of removing food reward. When these definitions are observed, Fig. 1 reveals considerable evidence consistent with the hypothesis that extinction trials are reinforcing.

There are two approaches one might use to test for reinforcing effect of extinction in these data. First, one might reason that if extinction trials are as reinforcing as rewarded trials, then all groups should perform equally well at the end of Day 9 because they have all had the same number of trials. One might then perform a statistical test on the data from the last trials of Day 9 and conclude the acceptability of the hypothesis (Cotton, 1957) that extinction is as reinforcing as acquisition if the test yields no significant superiority in speed for Group A relative to the other groups. Such a test has already been described in connection with the terminal level of performance. Not only was Group A not superior to the other groups; it was actually inferior. Thus one acceptable conclusion is that extinction is *at least* as reinforcing as acquisition and perhaps more reinforcing. Alternate interpretations are not excluded by the present design, however.

Another approach involves the comparison of the mean speeds of the 50% groups on the last rewarded trials of the training sequence with the mean speed of Group A at the point in training at which this group had as many rewarded trials as the 50% groups had at the end of training. At the end of Day 9 the 50% groups had received 150 rewards, the same number that Group A had received at the end of Day 5. Any superiority in speed of the 50% groups at the end of Day 9 relative to Group A at the end of Day 5 can be attributed to the experience of extinction trials by the 50% groups during Days 2 through 9. An analysis of variance was performed using the mean speed on the last 5 trials of Day 5 for Group A, the mean speed on the last 5 trials of Day 9 for Group R, and the mean speed on the last 5 rewarded trials of Day 9 for all other groups. An F of 3.80 was obtained which is significant beyond the .01 level (6/133 *df*). A subsequent Tukey analysis revealed that Group A was slower than all groups except Groups 1E-1A and 30E-30A, at the .05 level or better, and that Groups 1E-1A and 30E-30A were almost significantly faster than Group A. No other comparisons were significant. Thus the result of this test is consistent with the hypothesis that the extinction trials had a reinforcing effect and that the magnitude of the effect is probably not a function of the pattern of rewarded and nonrewarded trials.

Within-day decrement for Group A.—Figure 1 reveals that there is a growing tendency between Days 2 and 10 for the last trials of Group A (in a block of 30 trials) to be slower than the first trials. As a test of this trend within a day, a Kogan (1948) analysis was performed using six 5-trial means on Day 9 for each *S* in Group A. The F obtained is 3.38, which is highly significant ($P < .01$, 5/95 *df*), indicating a significant downward trend in mean speed. That this trend cannot reasonably be attributed to either reactive inhibition or food satiation (from the reward pellets) is indicated by the following result: Group 30E-30A, with equal massing throughout, shows no such downward trend in Fig. 1. When Group 30E-30A is included in the above-mentioned Kogan analysis, the F obtained for interaction of trials and

groups is significant at beyond the .03 level. Thus Group 30E-30A, which expended equal energy under equal massing of trials and received equal food reward relative to Group A, certainly doesn't show a parallel downward trend and presumably no downward trend at all.

It is of interest to note that the *F* between groups in the Kogan analysis was only 1.49 with 1 and 38 *df*, indicating that Groups 30E-30A and A did not differ significantly on *mean* speed (for 30 trials) during Day 9 even though a previous analysis had shown that Group A was significantly slower during the last 5 trials of Day 9.

Discussion

Acquisition and terminal level of performance.—The acquisition data confirm the findings of Weinstock (1958), Notterman (1951), Lauer and Carterette (1957), and Goodrich (1959) in that the 100% group (A) did not reach as high a terminal acquisition level as the 50% groups. The results of all five studies could be interpreted as evidence counter to Weinstock's (1954) explanation of partial reinforcement in terms of competing responses, since that theory predicts equal asymptotes for groups receiving different proportions of reinforcement. However, in the present study as in the other studies, the differences are not large; a theory predicting equal asymptotes provides a good approximation to the data. The mean speeds (for all 30 trials) of Groups A and 30E-30A for Day 9 were not significantly different (Kogan test) and very likely Group A did not differ significantly from any other group on overall performance during Day 9. This finding also limits the evidence counter to any theory which predicts equal asymptotes.

An intriguing feature of the acquisition data concerns the within-day decline in speed for Group A beginning on Day 3 and continuing through Day 9. It has already been argued in connection with the Kogan analysis performed on the scores for Day 9 that this decline cannot be explained in terms of either muscular fatigue (inhibition) or food satiation because of the control provided by Group 30E-30A. What, then, does explain the decline? The present data do not allow of any conclusion in this matter, but the *ad hoc* definition will be advanced that this is "boredom" behavior associated with the often-repeated routine of running to a uniform, unvarying reward. Perhaps a theory such as Glanzer's (1953) theory of stimulus satiation can be made to handle this phenomenon. It is of interest to note that the data of Lauer and Carterette (1957) show a related phenomenon. Their *S*s which received continuous reward reached a peak reciprocal latency on Trial 72; this score was never equalled or exceeded on the remaining 36 trials of the training sequence: on the contrary, most of the scores fell well below this peak value. Similarly, a reciprocal running time score of about .54 sec. was obtained on Trial 58, and this value was exceeded only once during the subsequent 50 trials of training. Since the intertrial interval was either 30 min. or 24 hr. in the Lauer and Carterette study and only about 30 sec. in the present study, the "boredom" phenomenon does not seem to be related to intertrial interval. If the phenomenon proves to be reliable, the implication is that habit strength cannot be fruitfully conceived to be a monotonic increasing function of number of rewarded trials, at least not in situations wherein a very extensive training sequence is employed.

In connection with asymptotic behavior, it should be noted from Fig. 1 that running speed for partially reinforced groups was still increasing after 240 trials (end of Day 8). Weinstock's data (1958, Fig. 1) seem to indicate that an asymptote was reached after only 90 trials. This discrepancy may mean that trial spacing is a variable related to rate of approach to the asymptote of maze performance.

Extinction.—The fact that all groups except Group 30E-30A showed partial reinforcement effects on Day 10 relative to Group A is consistent with the notion advanced by Lauer and Estes (1955) and by Razran (1955) that the partial reinforcement effect upon extinction is a special case of repeated acquisitions and extinctions. The failure of Group 30E-30A to exhibit partial reinforcement effects may mean that grouping of rewarded and nonrewarded trials will produce partial reinforcement effects only up to a certain point (less than 30 trials in a block). However, Group 30E-30A was the only group which did not experience rewarded trials on the same day that it experienced nonrewarded trials, and that fact opens the door to another interpretation. It is possible that "sizable" blocks of nonrewarded trials will not affect resistance to extinction unless they are followed in "relatively short" (less than 24 hr.) temporal contiguity by rewarded trials. Weinstock (1954, 1958) has on two occasions shown that giving a single nonrewarded trial on alternate days is sufficient to produce the partial reinforcement effect. There is no evidence, however, that larger blocks of nonrewarded trials administered on alternate days will produce the effect.

Reinforcing effects of extinction.—The fact that extinction trials during Days 1 through 9 produced decrements in running speed followed by increases to the level of the control group (Group A) or above when matched on total number of *trials* suggests that one aspect of the response system, hereafter identified with Hull's habit strength (Hull, 1943), is increasing on nonreinforced trials. This is consistent with Spence's (1956) position. Another aspect of the response system, identified with Hull's (1951) incentive motivation (K), appears to increase somewhat exponentially with introduction of reward and decline exponentially with removal of reward. This reciprocal action of K with reward and nonreward is necessary to prevent instantaneous shifts in running speed with magnitude of reinforcement changes (Crespi, 1942) and, as pointed out by Dufort and Kimble (1956), is then sufficient to account for extinction and reconditioning as well.

One might argue that secondarily reinforcing stimuli are present on extinction trials, producing the increments to reacquisition just noted. These stimuli might be associated with handling, with the goalbox, or with the activity of running. The question arises whether there are ever extinction trials in which such secondary reinforcements are completely absent. Until such cases are found, it seems reasonable to be contiguity theorists and join Spence (1956) and Guthrie (1935) in predicting learning whenever a response is executed, even though increments in behavior may not appear until drive or incentive motivation is later increased.

Summary

Rats were given 30 rewarded trials in a runway on Day 1. Five groups of 20 *Ss* each received alternating blocks of nonrewarded and rewarded trials during Days 2 through 9 (30 trials per day): the block sizes were 1, 5, 10, 15, and 30 trials. One group of 20 *Ss* received reward on a randomly assigned 50% of the 240 trials given during Days 2 through 9 and another group (Group A) received reward on all 240 trials. On Day 10 all groups received 30 nonrewarded trials.

Group A ran significantly slower than most other groups during the last 5 rewarded trials of Day 9, and the other groups did not differ significantly among themselves. Group A showed a significantly slower mean speed on Day 10 than all other groups except the groups which had received blocks of 10 and 30 nonrewarded trials. Group A ran significantly slower during the last 5 trials of Day 10 than all groups except the group which had received blocks of 30 nonrewarded trials.

The results support the assumption that nonrewarded trials increase response strength on subsequent reacquisition trials. The study also supports the conjecture that partial reinforcement effects are but a special case of repeated acquisitions and extinctions.

References

Cotton, J. W. The reinforcing effects of extinction. Paper read at the meetings of the Midwest. Psychol. Ass., 1957.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

DUFORT, R. H., & KIMBLE, G. A. Changes in response strength with changes in the amount of reinforcement. *J. exp. Psychol.*, 1956, **51**, 185–191.

GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JENKINS, W. C., & STANLEY, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, **47**, 1950, 193–234.

KOGAN, L. S. Analysis of variance-repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

LAUER, D. W., & CARTERETTE, T. S. Changes in response measures over repeated acquisitions and extinctions of a running habit. *J. comp. physiol. Psychol.*, 1957, **50**, 334–338.

LAUER, D. W., & ESTES, W. K. Successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcement. *J. comp. physiol. Psychol.*, 1955, **48**, 8–13.

LEWIS, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *J. exp. Psychol.*, 1956, **51**, 45–53.

NOTTERMAN, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *J. exp. Psychol.*, 1951, **42**, 273–291.

RAZRAN, G. Partial reinforcement of salivary CR's in adult human subjects: Preliminary study. *Psychol. Rep.*, 1955, **1**, 409–416.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *Amer. J. Psychol.*, 1953, **65**, 57–65.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received July 22, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 50-56

EFFECT OF SEQUENTIAL PATTERNS UPON SERIAL-POSITION ERRORS AND ACQUISITION OF A VERBAL MAZE [1]

GEDIMINAS NAMIKAS AND W. J. BROGDEN

University of Wisconsin

Limited modification of the 16-unit, 4-choice maze pattern of Brogden and Schmidt (1954) by the introduction of doublets and split doublets (Ernst, Hoffeld, Seidenstein, & Brogden, 1960; Thompson, Voss, & Brogden, 1957;) and of triplets and quadruplicates (Namikas, Thompson, & Brogden, 1960) alters the form of the serial-position curve for errors but does not affect speed of acquisition. Each of the different forms of sequence of correct responses for the maze pattern has a characteristic effect that is largely independent of locus within the maze. The phenomena involve both increases and decreases in error. Some vary as a function of stage of acquisition and type of error measure (total or first error) and the doublet phenomenon is related to the tendency of *S*s not to repeat a response correct for for the previous position.

More substantial changes in the form of the serial-position curve for errors and in speed of acquisition might be expected from sequential patterns that involve all positions of the maze. Within the limits of the 16-unit, 4-choice (10, 20, 30, and 40) maze in which each choice is represented four times, there are a number of sequences that might be used. Since it has not been possible to develop any scale for this kind of pattern variation, an experiment was conducted on the effect of three unique pattern sequences relative to the standard pattern of Brogden and Schmidt (1954). Each pattern was designed to increase the speed of acquisition required for the standard pattern by providing sequential regularity over the 16 positions, thus introducing some aspects of problem solving to the maze task. Although there were no specific hypotheses about the form of the serial-position curve for errors, each pattern was expected to produce a characteristic curve.

[1] This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

PROCEDURE

The first of the modified patterns was learned by Group E_1 and consists of the four alternate choices in repetitive ordinal sequence throughout the entire maze (10, 20, 30, 40, 10 . . .). This pattern was permuted to give three additional patterns, starting with 20, 30, and 40, and to provide that over the four subpatterns, each number appear once at each of the 16 positions. One-quarter of the *S*s in Group E_1 learned each of the four subpatterns.

For Group E_2, each number was presented as a doublet (10, 10, 20, 20, 30, 30, 40, 40, 10, 10 . . .). Permutation of this pattern provides four submazes, each of which was learned by one-quarter of the *S*s.

The pattern for Group E_3 involved successive quadruplicates in ordinal series (10, 10, 10, 10, 20 . . .) and again the four permutations provide four subgroups of the *S*s.

The *S*s were 240 volunteer students from the classes in elementary psychology at the University of Wisconsin. They were assigned randomly, in order of their appearance, to one of the four subgroups in each of the three main groups so that there were 20 in each subgroup and 80 in each main group.

The instructions presented by Thompson et al. (1957) were read to each *S*. The *E*

indicated each correct response made by *S* by saying "Forward." There was a rest interval of 1 min. after each trial. The rest interval and the duration of each trial were timed by two alternating electric timers. The length of each trial and every error for each position was recorded by *E* on prepared data sheets. Each *S* learned his maze pattern to the criterion of one errorless trial.

Results

Data from the control group reported by Ernst et al. (1960) were used to evaluate the present experimental findings. Analyses of variance were performed on time, trials, and total errors to the criterional trial for the three experimental groups and the control group. Because there is heterogeneity of variance for all measures, the 1% level of significance will be used for evaluation of these and all subsequent analyses. All three analyses are consistent in showing maze pattern and the interaction of maze pattern and submaze pattern to be significant sources in variation. Range tests (Duncan, 1951) on the time and trials data show the means for the experimental groups to be significantly smaller than the means for the control group, but not significantly different among themselves (for time, $E_1 = 73.7$ sec., $E_2 = 105.6$ sec., $E_3 = 93.0$ sec., $C = 772.7$ sec.; for trials, $E_1 = 2.4$, $E_2 = 3.2$, $E_3 = 2.5$, $C = 18.4$). The interaction of Maze Pattern and Submaze Pattern is shown most clearly by the error data, so that for time and trials is not presented.

For the error data, Serial Position, Maze Pattern × Serial Position, and

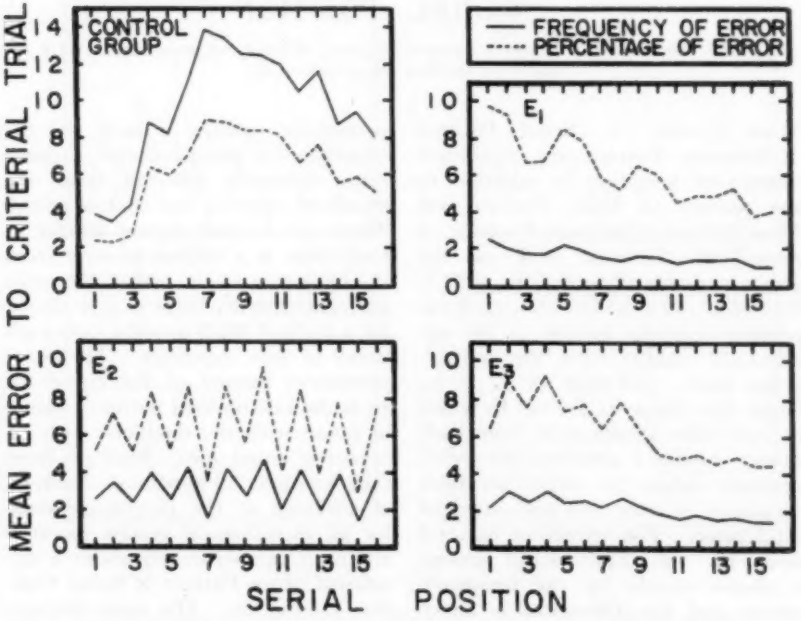

FIG. 1. Serial-position curves in terms of frequency of error and percentage of total errors for the main experimental and control patterns.

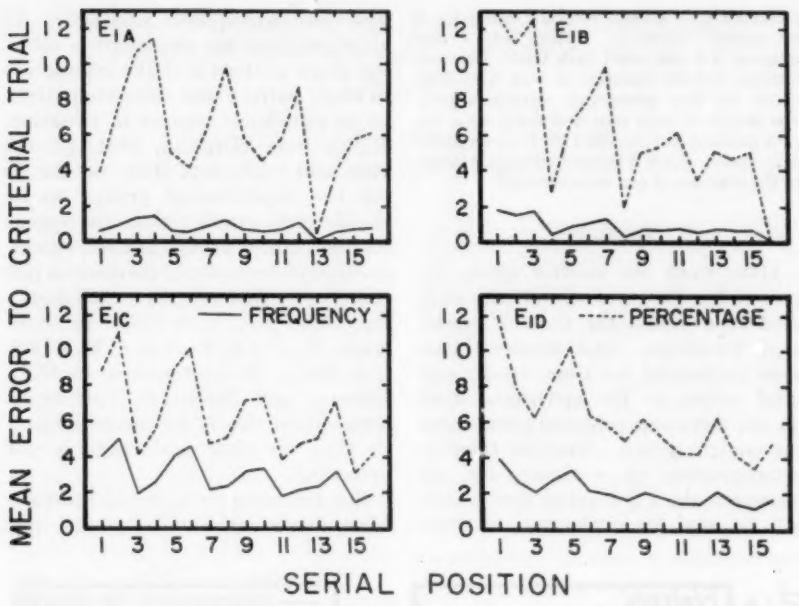

FIG. 2. Serial-position curves in terms of frequency of error and percentage of total errors for the four subpatterns of E_1.

Maze Pattern × Serial Position × Submaze Pattern are significant sources of variation in addition to the sources of Maze Pattern and Maze Pattern × Submaze Pattern. A range test (Duncan, 1951) on the group measures ($E_1 = 25.7$, $E_2 = 48.9$, $E_3 = 33.2$, $C = 156.9$) shows all experimental-group means to be significantly smaller than the control-group mean, and that for E_2 to be larger than those for E_1 and E_3 which do not differ significantly from each other. Figure 1 presents the serial-position curves for errors for both frequency of error and percentage of total error. The reduction in total error for the experimental groups is shown clearly by the frequency curves and the differences in curve form are shown most clearly by the percentage curves. Each of the experimental groups shows a curve form markedly different from the standard curve of the control group. Those for E_1 and E_3 are similar in that there is a decline in error from the beginning to the end of the maze pattern. For E_1 there is a tendency for a cyclical effect to occur over each series of four positions. The most prominent feature of the curve for E_2 is the saw-toothed pattern, cycling in phase with the doublets, with an apparent equal mean level of error otherwise over all positions. Analysis of variance of the percentage data for all experimental groups together and each pair separately shows a significant Maze Pattern × Serial Position interaction. The maze pattern by serial position by submaze pattern

is significant for comparisons of E_2 with E_1 and E_3, but not for the comparison of E_1 and E_3.

The serial-position curves for each submaze pattern are plotted in terms of both frequency and percentage of error for E_1 in Fig. 2, for E_2 in Fig. 3, and for E_3 in Fig. 4. Separate analyses of variance were completed for each group on both the frequency and percentage error data. For the frequency data, only in the case of E_1 is Submaze Pattern a significant source of variation. A range test (Duncan, 1951) shows significant differences between Pair A and B and Pair C and D, but not between the members of each pair (A = 12.2, B = 13.1, C = 45.2, D = 32.4). The frequency curves of Fig. 2 show these differences.

The interaction of Submaze Pattern and Serial Position is significant for each group for both the frequency and percentage error with one exception, E_1 for percentage error. The curves for E_1 all show similar cyclical variation with differences in phase between them. For E_2, the most striking difference is that between Submaze Pattern A and the other three submaze patterns. The cyclical pattern for the group as a whole seems to be largely a function of the pattern for Subgroup A. There is some cyclical effect for Subgroups B, C, and D, but in each of these curves there appear to be position effects also. The curves for submaze patterns of E_3 show some similarity in form for A and D, and for B and C, although there are also differences between the members of each pair.

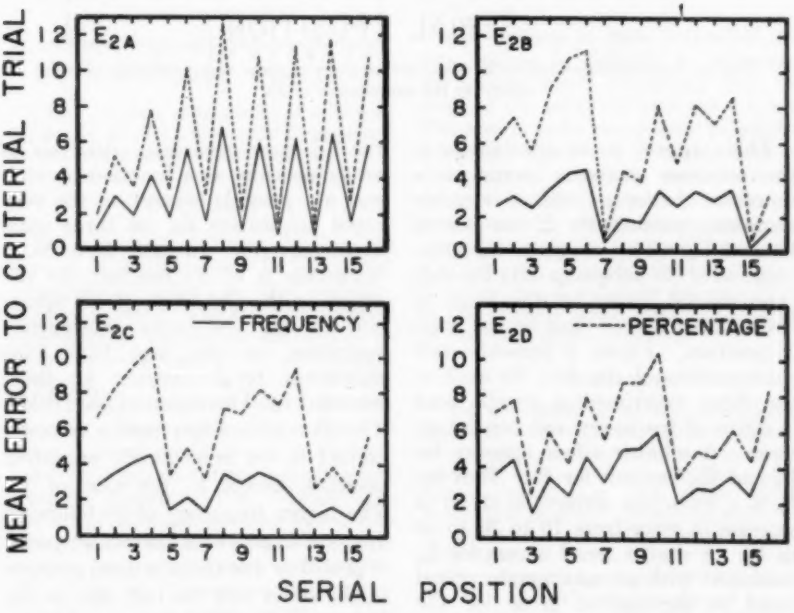

FIG. 3. Serial-position curves in terms of frequency of error and percentage of total errors for the four subpatterns of E_2.

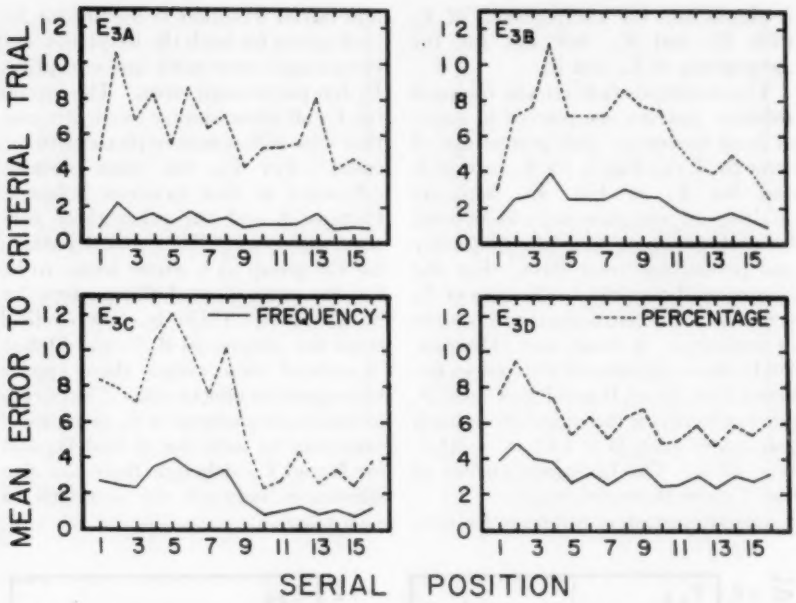

FIG. 4. Serial-position curves in terms of frequency of error and percentage of total errors for the subpatterns of E_3.

There appear to be regularities in the submaze pattern curves as a function of the individual member responses, particularly in the case of E_1 and E_2. This suggests the combination of the subgroup data for each experimental group on the basis of like number rather than on the basis of position. Figure 5 presents such combinations of the data for each of the three experimental groups both in terms of frequency and percentage error. A number effect appears for E_1 and E_2, but not for E_3. That for E_1 is a four-cycle sequential trend of increase in error from 10 to 20 to 30 to 40. A similar trend occurs for E_2 combined with an additional cyclical trend of increase in error for the second position of the doublet.

Analyses of the first errors gave similar results to those obtained for the all error measure. Analysis of either measure as a function of trials was not possible because of the very rapid acquisition for all three main maze patterns. In fact, three *S*s in Subgroup A of E_1 reached the criterion on the first trial. A chi square test of the first order contingency responses on the first trial gave significant results similar to those reported by Thompson et al. (1957). The *S*s tend not to repeat a response correct in the immediately preceding position, except in the case of 10. The higher frequency of 10 following 10 as the previous correct response is probably due to its ordinal position in the series and the fact that in the instructions, the alternate responses were presented in the order 10, 20, 30, and 40.

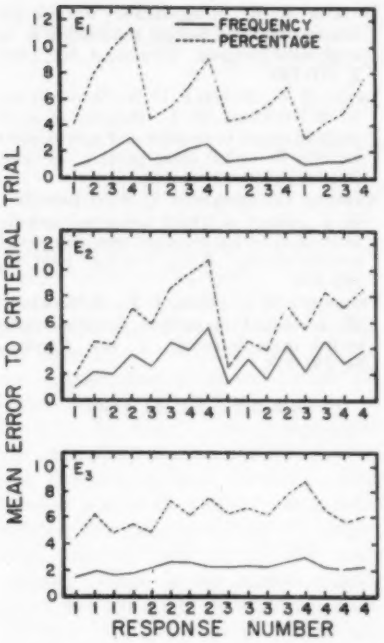

FIG. 5. Curves of frequency of error and percentage of total error for the three experimental groups based on like number.

DISCUSSION

The sequential patterns used in the present experiment result in very rapid acquisition relative to the standard random pattern. The repetitive ordinal sequence (E_1) and the successive quadruplicate in ordinal series (E_3) are of equal order of difficulty and less difficult than the pattern of successive doublets in two ordinal series (E_2). Each of these patterns produces a serial-position curve for errors that is markedly different from the standard curve. In each case, the effect of serial position appears to have been virtually eliminated, with relative frequency of error to be a function of the relations among the response members of the maze pattern. That the effect of pattern upon frequency of error is a function of the particular members is further shown by the significant differences between submaze patterns within each of the three main patterns. The recombination of the data on the basis of like response number shows the direct effect of number, particularly in the case of E_1 and E_3.

The relatively slower acquisition for E_2 than for E_1 and E_3 appears to be a function of the guessing habits of *Ss*, particularly the tendency not to repeat a response that was correct for the previous position. This results in an increase in error for the second item of each doublet. Such an effect will probably occur for any pattern involving doublets, whether a single doublet (Ernst et al., 1960; Thompson et al., 1957), repetitive doublets in ordinal series as in the present experiment, or other possible patterns of doublets. Repetitive sequential patterns that do not involve doublets should be learned relatively more rapidly and be at a level of difficulty comparable to that shown by E_1 and E_3.

The inability to scale sequential pattern makes it difficult to generalize the effects upon acquisition and the form of the serial position error curve. There seem to be two general principles, however. Relative to the random standard pattern, sequential patterns result in substantially faster acquisition and the standard serial-position curve for errors is completely destroyed by sequential patterns. The distribution of errors is a function of relations between items of the pattern rather than the position of the items in the sequence.

SUMMARY

The acquisition curves and serial-position curves for error of three sequential patterns involving all items of the 16-unit maze were compared with that for the standard 16-unit, 4-choice verbal maze. The pattern for E_1 involved a repetitive ordinal sequence (10, 20, 30, 40, 10 . . .), for E_2 successive doublets (10, 10, 20, 20, 30, 30, 40, 40, 10 . . .), and for E_3 successive quadruplicates (10, 10, 10, 10, 20 . . .). All measures showed acquisition of the sequential patterns to be significantly faster than acquisition of the standard pattern, and acquisition for E_1 and E_3 more rapid than for E_2. The slower acquisition

for E_2 appears due to the doublet effect of an increase in error for the second item of the doublet which in turn is a function of the guessing habits of *Ss* of not repeating a response correct for the previous position.

The serial-position curves for errors in the experimental groups are completely different from the standard curve. The form of the curve is a function of the relations of the numbers making up the pattern rather than their position. Recombination of the data on the basis of like number rather than position demonstrates this clearly.

References

BROGDEN, W. J., & SCHMIDT, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.

DUNCAN, R. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951. **2**, 171–189.

ERNST, R. L., HOFFELD, D. R., SEIDENSTEIN, S., & BROGDEN, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

NAMIKAS, G., THOMPSON, C. P., & BROGDEN, W. J. Effect of triplet and quadruplicate location in verbal maze pattern upon serial position errors. *J. exp. Psychol.*, 1960, **59**, 383–390.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, **54**, 253–258.

(Received July 23, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 57-63

THE ROLE OF VISUAL CUES IN THE PERFORMANCE OF ERGOGRAPHIC WORK [1]

LEONARD E. JARRARD [2]

Carnegie Institute of Technology

The intent of the present research was to provide a more precise specification of the effects of psychological factors in the activation of the human neuromuscular system. The specific problem centered around determining the extent to which muscular tension and ergographic performance are affected by set toward the amount of work to be done. The independent variable of set was varied by employing the size-weight illusion. This illusion is such that objects of greater volume are judged lighter in weight than smaller objects of the same weight. Thus, in the present experiment one group of *S*s changed from lifting a large- to a small-size weight that physically weighed the same. Such a change was perceived by *S*s as being a change to a heavier weight. A second group of *S*s changed from a small- to a large-size weight that was perceived lighter in physical weight. It was assumed that these visual cues would serve to establish a set toward the amount of work to be done.

Several investigators (Bills & Brown, 1929; Freeman & Lindley, 1931; Krueger, 1937) have found that set toward the amount of work to be done is important in affecting subsequent behavior. Initial levels of restlessness and tonus were found by Freeman and Lindley (1931) to be highest for *S*s that expected to work the shortest period of time. Krueger (1937) reported that the greater the amount of work to be done, the less was the early output in a number adding task. In these experiments, the set was established by giving *S* instructions that related either to the length of time the task was to be performed or the amount of work that was to be carried out. Such instructions evidently serve to affect *S*s' perception of the task and thus his performance.

It is hypothesized in the present study that those *S*s who change to the weight that they perceive heavier will lift with higher muscular tension. As a result, there will be a decrease in the number of lifts that can be made. For *S*s who change to the weight that they perceive as lighter, it is hypothesized that there will be a resulting decrease in muscular tension and thus they will be able to carry out a greater number of lifts.

METHOD

Experimental design.—There were four control and two experimental conditions. Data for three of the conditions were ob-

[1] This paper is based upon a dissertation submitted to the Department of Psychology at the Carnegie Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to Lee W. Gregg for his guidance and encouragement throughout the investigation.

The author was a Public Health Service Fellow of the National Institute of Mental Health during the time that this research was carried out. The project was also supported in part by Lee W. Gregg's research grant, G-2105, from the National Science Foundation.

This study was reported at the 1959 Midwestern Psychological Association meetings in Chicago.

[2] Now at Washington and Lee University, Lexington, Virginia.

tained two months after that for the other three. The conditions for the first half of the experiment were as follows: (*a*) Condition NVC was a control condition of no visual cues in which *Ss* could not see the weight that was being lifted. (*b*) The *Ss* in the control constant visual cues Cond. LLLL saw that they were lifting the same large-size weight throughout the session. (*c*) Condition LSSS was an experimental condition of altered visual cues in which *Ss* changed after preliminary experience with the large-size weight to a weight one-eighth the size that actually weighed the same. The conditions for the second half of the experiment consisted of the following: (*a*) Control Condition NVC was the same as above in that *Ss* could not see the weight. (*b*) Condition SSSS was a control condition in which *Ss* saw that they were lifting the same small-size weight throughout. (*c*) The *Ss* in experimental Cond. SLLL changed from the small- to the large-size weight. Thus, *Ss* in experimental Cond. LSSS changed to a weight that was perceived heavier and *Ss* in experimental Cond. SLLL changed to a weight that was perceived lighter.

Apparatus.—A modified finger ergograph was constructed which differed from the standard apparatus in that *S* could clearly see the weight he was lifting, and a hairline target indicated the height to which the weight was to be lifted.

The main part of the ergograph was housed in a wooden frame 21 in. high by 15 in. long by 14 in. wide. Extending out from the front part of the frame was an arm rest 28 in. long and 5.5 in. wide and tilted at a 5° angle. A system of four pulleys was arranged so that a nylon cable extended from a leather finger strap to the back of the frame, up and across the top, and then down to the front part of the apparatus. The cable was attached in front to an elevator-like platform upon which a weight could be placed. The entire front part of the ergograph had a plywood covering but the bottom half was removable; thus, *E* could permit *S* to see the weight he was lifting or this information could be withheld.

A second major modification consisted of a hairline target located midway in the upper half of the front of the ergograph. In the 1.5 in. by .75 in. aperture there were two wires .5 in. in length that extended horizontally from the midpoint of either side. A third wire was located horizontally on the movable elevator-like platform and was visible only when *S* lifted the weight above a certain height. With a 1000-gm. weight (as was used throughout the experiment) a lift of 43 cm. was necessary to perfectly align the wires. Recordings of each lift were obtained with a single-pen kymograph.

The wooden blocks used in lifting were of two sizes. A single, large-size, 1000-gm. weight was 15 cm. high with a square base 10 cm. by 10 cm. A series of seven, small-size weights were all 7.5 cm. high with a base 5 cm. square. The smaller weights had a volume one-eighth that of the larger one. These small-size blocks were drilled and some of them evenly filled with lead to adjust for weight. Actual weights for the seven blocks were 250, 375, 500, 625, 750, 875, and 1000 gm.

The electronic system that was used to measure and record the muscle action potentials (MAPs) has been described by Gregg (1958) and Gregg and Jarrard (1958). Basically, it consisted of seven distinct units: a preamplifier, a scanner, main amplifier, an oscilloscope, a decade-type electronic interval timer, a multiple integrator unit and the power supply.

The system was so arranged that MAPs from five separate muscle groups were amplified, sent through a double scanner, monitored by the oscilloscope, and fed into five integrators from which quantitative readings in meter microamperes were obtained. The interval timer served to control the period of scanning and provided audible clicks which indicated the beginning and end of a recording trial. For both experiments, the timer was adjusted so that the scanner went through two complete cycles or a 24-sec. scanning period. Meter units (mu.) are thus in terms of meter microamperes per two cycles.

Pairs of silver electrodes .5 in. in diameter were placed over five different muscle groups and held in place with Ace elastic bandages. The sequence in which the electrode pairs were scanned was controlled by a 10-pole, 5-gang selector switch. "The principle of the latin square was used in wiring this switch so that over a block of five trials each electrode pair was scanned in each ordinal position in time during the scanning cycle and the amplified potentials were recorded from each integrator unit" (Gregg & Jarrard, 1958, p. 532).

Subjects.—The *Ss* were 72 undergraduate men from introductory courses in psychology at Carnegie Institute of Technology. These volunteers participated in the study in order to complete a course requirement that they serve as *Ss* in at least 3 hr. of experimental work. Thirty-six *Ss* were randomly assigned to three conditions in each half of the experiment.

Procedure.—On entering the experimental room, *S* was asked to remove his clothing above the waist and have a seat inside the 6 ft. wide by 10 ft. long by 7 ft. high electrically shielded cage. The general nature of electromyographic recordings was explained as the five pairs of electrodes were being placed over five different muscle groups. Electrodes were placed over the *flexor digitorum sublimis, palmaris longus, flexor carpi radialis,* and the *biceps brachii* of the preferred arm. The trapezius from the back of the neck was recorded for all *S*s.

The *S* was told that the purpose of the experiment was ". . . to determine how accurately college students can lift different weights over a 40-min. period." Before starting work on the ergograph, *S* was asked to compare the large-size weight with each of the seven small weights and to pick out the small weight that weighed the same as the large one. The purpose of this comparison was to have *S* experience the size-weight illusion and to provide *E* with an indication of the extent to which *S* was affected by the illusion.

A block of five recording trials was then carried out while *S* relaxed. By changing the selector switch after each trial, each muscle group was recorded from each meter during a block of trials. Each trial lasted 24 sec., the period of two complete cycles of the scanner. Thus, during a single trial MAPs from each muscle group were sampled at two different times for a period of 2 sec. The 4-sec. difference between total time and the actual recording time was used in switching from one muscle group to another.

Following the resting block of trials, *E* demonstrated how to perform on the finger ergograph. Special emphasis was placed on keeping the palm up and flat, the finger stiff, and trying not to use the index and ring fingers. The *S* was told to watch the hairline target while lifting and to "align the wires" as near as possible on each lift. Ten practice lifts were then made by *S*.

No visual cues relating to size of the weight were available to *S*s in Cond. NVC. In the first block of lifts, Block 1, *S*s in Cond. LLLL and LSSS lifted the large-size weight; *S*s in Cond. SSSS and SLLL lifted the small-size weight during this first block of lifts. The proper rhythm was established by *E* calling cadence. This rhythm was such that *S* was lifting just as MAPs from a muscle group were being fed into an integrator. Thus, during one trial *S* lifted 10 times over the 24 sec. Between each trial there was a 48-sec. interval during which *S* relaxed while *E* recorded the integrated MAPs, changed the multiple-selector switch, and cleared the meters for another trial. After Trial 4 of the first block, *E* gave the following instructions:

> You will start the same as you have in previous trials but rather than stop after 10 lifts, I want you to continue lifting until you can't lift anymore—this will happen after your muscles tighten up and you lose control. It is especially important that you remember how your muscles feel when you stop because I will want you to lift several more times until your muscles feel the same way.

Following Trial 5 and this work period to exhaustion, *S* was permitted to remove his finger from the leather strap and rest for 2 min.

Next, Blocks 2, 3, and 4 were carried out with the only difference in procedure being a change in size of the weights for the experimental conditions. Thus, *S*s in Cond. LSSS changed from the large-size weight lifted in Block 1 to the small-size weight, while *S*s in Cond. SLLL changed from the small- to the large-size weight. The *S*s in all other conditions continued lifting the same size weights that they had in Block 1.

After the last block of lifts, the electrodes were removed and *S* was once again asked to compare the weights and to choose the small-size weight that weighed the same as the large one.

Results

Muscle action potentials.—For each of the 72 *S*s, a mean integrated potential value for the resting block of five trials was obtained by summing over muscle groups. These values were subtracted from similarly obtained means for each of the four subsequent blocks of trials. The resulting 288 scores consisted of 4 scores contributed by each *S*.

Block 1 served primarily to familiarize *S*s with the procedure and to provide practice on the task. Since the data in this block were obtained prior to the introduction of the experimental variable (change in size of the weights for the experimental conditions), these Block 1 data were

TABLE 1

MEANS AND *SD*s OF INTEGRATED MAPs AND PERCENTAGE OF NUMBER OF LIFTS TO EXHAUSTION

Cond.	*N*	MAP		Lifts to Exhaustion	
		Mean	*SD*	Mean	*SD*
NVC	24	21.2	3.7	68.8	26.0
LLLL	12	20.0	2.5	70.9	25.5
SSSS	12	21.7	2.6	70.7	20.6
LSSS	12	24.7	3.9	46.6	18.1
SLLL	12	20.7	3.5	107.2	38.2

analyzed separately. An analysis of variance indicated that there were no significant differences between conditions ($.10 < P < .20$). The groups were thus considered to be homogeneous with respect to the integrated MAPs produced in the lifting.

Table 1 presents means and *SD*s based on MAP differences between Block 1 and a mean value obtained by summing over Blocks 2, 3, and 4. A constant value of 20 was added to eliminate negative values. The two control Cond. NVC ($N = 12$) were combined because of similarity of means (20.8 and 21.5) and *SD*s (3.8 and 3.7) to form a combined Cond. NVC ($N = 24$).

An analysis of variance of the MAP data for the five conditions shown in Table 1 was carried out. The design of the analysis can best be characterized as a simple-randomized design with unequal *N*s (Lindquist, 1953). A significant Conditions effect ($F = 3.19$, $df = 4/67$, $P < .05$) indicates differences between conditions in terms of integrated MAPs. Differences in means for the individual pairs of conditions were tested for significance by applying the t test. The critical comparisons are those between the conditions receiving altered visual cues (LSSS and SLLL) and the control conditions (NVC, LLLL, and SSSS). The mean for Cond. LSSS was significantly higher than those for the other conditions ($P < .05$) indicating that the effect of changing from the large-size weight to the small one was to cause significant increases in integrated potentials for that condition. No significant differences were found between the means for any of the other conditions.

The effects of Muscles and Blocks were not evaluated in the above analysis. However, a separate, more complete analysis of variance showed significant differences between muscle groups and significant increases in integrated MAPs over blocks of trials. Further, a significant Conditions × Muscles interaction indicated a differential effect for muscles over the different conditions.

Performance measures.—For the height of lift data an analysis of variance of differences between the Block 1 mean and a mean for the three subsequent blocks was carried out with a design similar to that used for the MAP data. The results indicated that there were no significant differences between conditions in the height to which the weight was lifted.

The raw data for number of lifts to exhaustion consisted of mean percentages computed by dividing the number of lifts following Blocks 2, 3, and 4 by the number of lifts following Block 1. Table 1 presents the means and *SD*s for these data. A combined Cond. NVC ($N = 24$) was obtained by summing over the two control conditions. The requirement of homogeneity of variance was satisfied and an analysis of variance indicated that there was a significant difference between conditions ($F = 8.30$, $df = 4/67$, $P < .05$). Individual comparisons between conditions showed that *S*s

in Cond. LSSS lifted the weight significantly fewer times than control *Ss* and *Ss* in Cond. SLLL lifted a significantly greater number of lifts than the controls ($P < .05$).

The above analyses did not evaluate changes in performance over time (Blocks). However, a more detailed analysis of variance showed that there were no changes during the session either in the height to which the weights were lifted or in the number of lifts made following the blocks of trials.

Judgment of weights.—When *Ss* were first asked to pick one of the seven small-size weights that weighed the same as the large, 1000-gm. weight, the most often chosen weight was the one weighing 625 gm. Twenty-two *Ss* changed their judgments (16 judged heavier and 6 lighter) at the end of the session but the 625-gm. weight still represented the mode. In neither judgmental period was the small 1000-gm. weight judged equal to the large 1000-gm. weight.

Discussion

It is apparent from the results that set toward the amount of work to be done was important in affecting subsequent behavior. This statement is substantiated by two major results. First, muscular tension was significantly higher than control *Ss* in those *Ss* who changed to the weight that they had previously judged heavier. The reverse did not hold for *Ss* who changed to the "lighter" weight. Second, there was a direct relationship between set toward the amount of work to be done and work output, i.e., those *Ss* who changed to the "heavier" weight lifted significantly fewer times to exhaustion than control *Ss*, while those *Ss* who changed to the weight that they had judged lighter lifted significantly more times than the controls.

The fact that set toward the amount of work to be done tended to cause peripheral changes in one direction (increased muscular tension) but not the other is of considerable interest. For those *Ss* who changed to the weight that was perceived as heavier, there was an increase in the integrated MAP values thus indicating an increase in the rate of contraction of previously inactive motor units (Darcus, 1953). Similar results are found when either rate of lift or magnitude of load are increased (Klein, 1951). There was no corresponding decrease in integrated MAPs for those *Ss* who changed to the weight that was perceived as lighter. It seems reasonable to assume that under "normal" conditions of lifting and after practice on the ergographic task, an individual will lift in such a way that the least effort is required, i.e., a minimum number of motor units need be activated. Actually, this ideal is probably seldom realized but it must at least be approximated. It seems unlikely that one could perform the same amount of work by employing a significantly fewer number of motor units than he uses under so-called normal conditions. Thus, for the present study, it is possible that the experimental procedure limited the direction in which peripheral changes could take place.

The present findings agree with those of previous investigators who have reported differences in performance due to set toward the amount of work to be done (Bills & Brown, 1929; Freeman & Lindley, 1931; Krueger, 1937). It is not surprising that the group of *Ss* exhibiting the highest muscular tension lifted the weight a significantly fewer number of times to exhaustion than the control groups. Since *Ss* who changed from the small- to the large-size weight did not differ from control *Ss* in muscular tension, it would seem to follow that these groups would not differ in work output. That this was not the case is indicated by the significant differences between the groups in number of lifts

to exhaustion. It is felt that these results demonstrate the important role that the central nervous system plays in influencing the amount of work that will be done before the point of subjective exhaustion is reached.

The ergographic performance measures of height of lift and variability did not indicate any differences that could be attributed to set. Thus, all groups lifted the same weight over the same distance during the time that the MAPs were recorded, but the group that perceived the weight as heavier did this with higher muscular tension than the other groups. In terms of efficiency of performance, those *Ss* who lifted the "heavier" weight worked less efficiently than control *Ss* since the same amount of work was performed but with greater physiological "cost" to the individuals.

A comparison of the physiological and behavioral changes that resulted from the present experimental manipulation of set is of theoretical importance. The question of central or peripheral locus of set has been with us since the days of Titchener and in a 1941 review of the problem, Gibson (1941) concluded that the question "remains unsolved." In the present experiment there were peripheral changes that accompanied one condition of set but the procedure was such that peripheral changes were not recorded when an opposite set was given. That set was still operating but primarily at a "central" level is suggested by the performance measures for those *Ss* in which no peripheral changes were recorded. For the conditions used in the present study, set was demonstrated to affect peripheral changes and performance in one condition and only to affect performance in another.

Summary

This experiment was carried out to determine the extent to which muscular tension and ergographic performance are affected by set toward the amount of work to be done. The independent variable of set was varied by employing the size-weight illusion in such a way that *Ss* in the experimental conditions changed from lifting one weight to a second weight that physically weighed the same but differed in size (smaller or larger). Due to the effect of the size-weight illusion, such a change in size had previously been judged by *Ss* as a change to a heavier or lighter weight. The *Ss* in one control condition could not see the weight they were lifting. The *Ss* in two other control conditions could see they were lifting the same weight (small or large) throughout the session.

The principle results were as follows:

1. Muscular tension was significantly higher than control *Ss* in those *Ss* who changed to the weight that they had previously judged heavier. The reverse did not hold for *Ss* who changed to the weight that they had judged lighter.

2. There was a direct relationship between set toward the amount of work to be done and work output as indicated by number of lifts to exhaustion, i.e., those *Ss* who changed to the "heavier" weight lifted significantly fewer times than control *Ss* while those *Ss* who changed to the "lighter" weight lifted significantly more times than the controls.

Thus, there were peripheral changes that accompanied one condition of set but the procedure was such that peripheral changes were not recorded when an opposite set was given. That set was still operating but primarily at a "central" level is suggested by the performance measures obtained for the condition in which no peripheral changes were recorded.

References

Bills, A. G., & Brown, C. B. The quantitative set. *J. exp. Psychol.*, 1929, **12**, 301–323.

Darcus, H. D. Some effects of prolonged muscular exertion. In W. F. Floyd & A. T. Welford (Ed.), *Fatigue*. London: Lewis, 1953. Pp. 59–68.

Freeman, G. L., & Lindley, S. B. Two neuromuscular indices of mental fatigue. *J. exp. Psychol.*, 1931, **14**, 567–605.

Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.

Gregg, L. W. Changes in distribution of muscular tension during psychomotor performance. *J. exp. Psychol.*, 1958, **56**, 70–77.

GREGG, L. W., & JARRARD, L. E. Changes in muscle action potentials during prolonged work. *J. comp. physiol. Psychol.*, 1958, **51**, 532–535.

KLEIN, S. J. Muscle action potentials produced in various ways and their relation to quantitative and qualitative measures of ergographic work. Unpublished doctoral dissertation, Univer. of Pennsylvania, 1951.

KRUEGER, W. C. The influence of amount limits upon the rate of work. *J. appl. Psychol.*, 1937, **21**, 113–118.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received July 23, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 1, 64

REPLICATION REPORT: ALTERNATION FOLLOWING EXPOSURE WITHOUT CHOICE[1]

WILLIAM N. DEMBER

University of Cincinnati

Walker, Dember, Earl, Fliege, and Karoly (1955) were unable to find alternation behavior following passive exposure of rats to one goalbox of a T maze; Glanzer (1958) has recently reported positive results in such a situation. The present experiment had a dual purpose: to see whether Glanzer's results could be replicated and to test the suggestion made by Dember and Fowler (1958) that accessibility of the choice-point region is necessary if positive results are to be obtained.

Method.—The *Ss* were 96 male, albino rats, approximately 90 days old at the start of the experiment. All *Ss* had had eight nonrewarded trials in the experimental apparatus over the three days prior to the experiment, making their familiarity with the maze similar to that of the Walker et al. *Ss*. The *Ss* were individually caged, and kept on an ad libitum diet of Purina chow and water. The experiment was conducted in six replications of 16 *Ss* each, over a period of three weeks.

The maze (width = 4 in.; wall height = 4.5 in.) had a gray startbox 8 in. long, a gray stem 12 in. long, one black goal arm and one white arm, each 20 in. long. A partition could be inserted in the middle of either goal arm to create a compartment *adjacent* to the choice point and one *far* from the choice point. The partitions were identical in brightness, texture, and size with the guillotine doors at the choice point and with the panel at the far end of each arm.

Two main groups of 48 *Ss* each were formed. The *Ss* of one group were individually placed, for 1 min., in the adjacent compartment; the *Ss* of the other group were placed in the far compartment. Half the *Ss* of each group were exposed to the black arm, half to the white. Within 30 sec. following exposure, each *S* was given a free trial, with the partitions removed and the guillotine doors up.

Results.—The data are given in Table 1. Of the 96 *Ss*, 64 entered the nonexposed arm on the free trial. The number of *Ss* alternating is significantly greater than the 48 expected by chance ($\chi^2 = 10.67$, $df = 1$, $P < .01$). There was no significant difference, however, between *Ss* in the adjacent and the far groups, with 31 of the former alternating and 33 of the latter. Each of the two groups alone showed a significant frequency of alternation. For the adjacent group, $\chi^2 = 4.00$, $P < .05$; for the far group, $\chi^2 = 6.75$, $P < .01$. A nonsignificant ($\chi^2 = 3.00$, $.10 > P > .05$) preference for the black arm is revealed in the table.

The over-all results are in agreement with Glanzer's. The suggestion that locus of confinement could account for the difference in results between the Walker et al. study and Glanzer's was not confirmed.

TABLE 1

NUMBER OF *Ss* ALTERNATING AND REPEATING IN EACH EXPOSURE CONDITION

Compartment	Arm	Number of *Ss*		Total *N*
		Alternating	Repeating	
Adjacent	Black	14	10	24
	White	17	7	24
Far	Black	14	10	24
	White	19	5	24
Total		64	32	96

[1] This research was supported by a grant from the National Science Foundation. The animals were run by Eileen Beier.

REFERENCES

DEMBER, W. N., & FOWLER, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.
GLANZER, M. Stimulus satiation in situations without choice. *J. comp. physiol. Psychol.*, 1958, **51**, 332–335.
WALKER, E. L., DEMBER, W. N., EARL, R. W., FLIEGE, S. W., & KAROLY, A. J. Choice alternation: II. Exposure to a stimulus or stimulus and place without choice. *J. comp. physiol. Psychol.*, 1955, **48**, 24–28.

(Received January 27, 1960)

PSYCHOLOGICAL MONOGRAPHS

Coming in September . . .

No. 497. **A Re-evaluation of the Meaningfulness of Possible CVC Trigrams**

By E. JAMES ARCHER Price $1.50

In this study association values for 2480 trigrams (nonsense syllables) were determined, and two measures of reliability were obtained.

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION, 1333 16th St. N.W. Washington 6, D. C.

PSYCHOLOGICAL MONOGRAPHS

No. 484. **Impaired Performance with Acute Sleep Loss**

By H. L. Williams, Ardie Lubin, and Jacqueline J. Goodnow

Price: $1.00

No. 490. **Perceptual Conditions of Association**

By S. E. Asch, John Ceraso, and Walter Heimer

Price: $1.50

No. 491. **Meaning and Familiarity in Associative Learning**

By William Epstein, Irvin Rock, and Carl Zuckerman

Price: $1.00

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Dept. JEP
1333 16th Street, N.W.
Washington 6, D. C.